AFIT/GIR/LAS/99D-3

POWER, PERFORMANCE, AND PERCEPTION (P³):
INTEGRATING USABILITY METRICS AND
TECHNOLOGY ACCEPTANCE DETERMINANTS TO
VALIDATE A NEW MODEL FOR PREDICTING
SYSTEM USAGE

THESIS

Alan P. Fiorello, Captain, USAF

AFIT/GIR/LAS/99D-3

20000307 069

POWER, PERFORMANCE, AND PERCEPTION (P³):

INTEGRATING USABILITY METRICS AND TECHNOLOGY ACCEPTANCE

DETERMINANTS TO VALIDATE A NEW MODEL

FOR PREDICTING SYSTEM USAGE

THESIS

Presented to the Graduate School of Engineering and Management

of the Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the Degree of Masters of Science in

Information Resources Management

Alan P. Fiorello, BA

Captain, USAF

December 1999

AFIT/GIR/LAS/99D-3

POWER, PERFORMANCE, AND PERCEPTION (P^3):

INTEGRATING USABILITY METRICS AND TECHNOLOGY ACCEPTANCE

DETERMINANTS TO VALIDATE A NEW MODEL

FOR PREDICTING SYSTEM USAGE
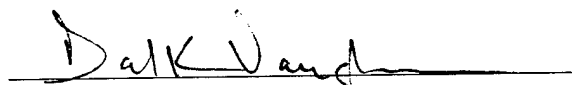
Alan P. Fiorello, B.A.
Captain, USAF

Approved:

_____

Michael G. Morris, Major, USAF, Ph.D.

Advisor

9 DEC 99
date

_____

Dr. D. Kirk Vaughan, Ph. D.

Reader

13 Dec 99
date

Acknowledgements

First of all, I must thank God for giving me strength and patience, and for teaching me humility during this thesis process. Secondly, I am eternally indebted to my wonderful and loving wife Judy for single handedly raising our children, running the household, and not leaving me during this entire degree program. Furthermore, I must deeply apologize to my two unconditionally loving sons, Jake and Connor, for not showing up for them as a Dad as my focus on this thesis resulted in their neglect.

I must also thank my "partner in crime," 1Lt Sean Abell, whose faith, dedication, professionalism, and intellect immeasurably contributed to the successful completion of this work. I could NOT have done it without you Sean, thanks again! I would also like to profoundly thank my thesis advisor, Major Michael G. Morris, for his guidance, insights, motivation, and commitment, which enormously enriched my learning experience during the thesis process and my entire time at AFIT. HCI Rules! Additionally, to my reader, Dr. D. Kirk Vaughan, your objective feedback was "dead-on" and I appreciate your patience with my less than stellar writing ability.

Finally, I would be remiss if I did not acknowledge the other individuals whose inspiration and support where instrumental in my completion of this thesis. An emphatic thank you to Mom, Pop and Norma, Major David Biros, Mr. Ovrin Shamblin, SSgt. Kelly Schell, Dr. Alan Hemminger, Lieutenant Colonels Lee Colburn and Evelyn Leahey, Dr. Pittman and his staff, the GIR/GIS 11th hour thesis writing team, Captains Barry Krauss and Greg Sears, Rudy "Never Quit" Reuttiger, and my loyal dog Solomon.

<div align="right">Alan P. Fiorello</div>

Table of Contents

List of Figures

List of Tables

Abstract

Currently, there are two distinct approaches to assist information technology managers in the successful implementation of office automation software. The first approach resides within the field of usability engineering, while the second approach is derived from the discipline of management information systems (MIS). The usability engineering approach has focused the question, "*can* users use the system?" while the MIS approach has answered the question, "*will* users use the system?" However, neither approach has successfully produced conclusive evidence that explains what characteristics facilitate system use as well as influence user acceptance of the system.

This study reports on the validity of a new model, entitled the Power, Performance, Perception (P^3) model, that links the constructs of usability engineering to user acceptance. For this study, speech recognition software (SRS), selected as the target technology due to its novelty and practical application to office automation software, was used in an experimental setting to validate the P^3 model. As a secondary focus, this research also examined the viability of employing SRS in an Air Force office setting.

The results of this study failed to validate the P^3 model. However, an alternate model for predicting user acceptance, the Usability-Acceptance Model, did emerge from the research which showed that the usability metric of user satisfaction can explain 53% of the variance of user intention to use a new technology. Additionally, the results of this study indicate that while users in a typical Air Force office environment would utilize SRS for text processing, the issue of increased productivity bears further examination.

Power, Performance, and Perception (P³):

Integrating Usability Metrics and Technology Acceptance Determinants to

Validate a New Model for Predicting System Usage

I.   Introduction

Background

Information technology has become integral, if not essential, to the operation of

modern-day organizations. Corporations are investing large amounts of money in

information technology by purchasing business information systems not only to maintain

a competitive edge but to simply remain in business. For example, companies in the

insurance and credit industry would not exist without aid of information systems (Laudon

and Laudon, 1997:7). Additionally, information technology accounts for 70 percent of

invested capital in service industries such as finance, insurance, and real-estate (Laudon

and Laudon, 1997:4).

Today's office information systems typically consist of networked desktop

computer systems that operate office software suites containing applications for word

processing, spreadsheet analysis, graphics, and e-mail, as well as a corporate database

management system. These systems were designed and built to act as assistants to office

workers to increase worker productivity and ultimately increase corporate profits

(Landauer, 1997:7). However, past studies have indicated that in the area of office

automation, the business community has experienced marginal gains at best in worker productivity compared to the dollars spent on information systems (Brynjolfsson and Hitt, 1993, 1996; Franke, 1987; Loveman, 1986, 1990). Therefore, it is vital that these investments in information systems be effectively and efficiently utilized by the acquiring organization to justify these costly investments.

Advances in the capabilities of computer hardware are occurring at an astonishing rate. According to Moore's Law, microprocessor speeds tend to double every 18 months (Horn, 1998; Intel, 1999), while the costs of personal computers have stayed relatively constant over the years and have recently been on the decline. These advances in computer hardware have yielded desktop systems with an incredible amount of processing power. Thus, computer hardware does not seem to be a main contributor of this computer productivity paradox as first identified by Brynjolfsson (1993). Therefore, by the process of elimination, computer software is likely to be a major contributor to this productivity problem. Landauer (1997:7) states that the problem with regard to computer software is threefold: 1) it is still too hard operate, 2) applications get misused, badly applied, and to wrong jobs, and 3) the most significant problem is that software applications yet do not do a sufficient number of significantly useful things. These shortcomings of computer software identified by Landauer are generally categorized as either *usability* or *usefulness* problems. The difference between usability and usefulness is not always clear (Landauer, 1997:143); however, each term does have a distinct meaning with respect to software. Usability focuses on how well users can utilize the functionality of the software to accomplish a specific task or goal (Shackel, 1991:124; Nielsen, 1993:27). Conversely, usefulness is primarily concerned weather and the

2

software can be used to achieve some desired task or goal regardless of the effort of the user (Nielsen, 1993:24).

Ultimately, the end-user wields the greatest power in determining the success of the information system at the workplace. Investments in information technology incur the risk that users may reject the system and not use it regardless of the utility of the system (Davis and others, 1989:982). People resist adoption of new information technology due to their attitudes toward technology, past experiences with technology, poor system design, and the lack of system usefulness as it applies to accomplishing tasks in their organization (Markus, 1983:68). For example, an organization can spend thousands of dollars on the purchase of new office automation software, with the intention of improving worker productivity, to regretfully discover that the new software is difficult to use, that it is unreliable, and that the workforce prefers using the old manual method. Therefore, it is important that information technology managers understand a process of user acceptance before making major investments in office automation technology.

Currently, there are two distinct approaches to assist information technology managers in the successful implementation of office automation software. Each approach has its own merits; however, no single approach fully encompasses all the factors necessary to successfully deploy a new software system in an organization. The first approach resides within the field of usability engineering, while the second approach is derived from the discipline of management information systems (MIS).

The Usability Engineering Approach

Usability engineering is firmly rooted in the discipline of human computer interaction (HCI). HCI is a multidisciplinary field that encompasses researchers and practitioners with varying backgrounds in computer science, business, human factors, and cognitive psychology. Although there are many definitions for HCI, the following seems the most concise and complete: "Human-computer interaction (or HCI) is, put simply, the study of people, computer technology and the ways these influence each other. We study HCI to determine how we can make this computer technology more usable by people" (Dix and others, 1993:xiii).

The term *usability* was intended to replace the term "user friendly," which by the early 1980s had acquired a host of undesirably vague and subjective connotations (Bevan and others, 1991). Usability, defined in the context of HCI, is the quality of a system that makes it easy to learn, efficient to use, easy to remember, error tolerant, and subjectively pleasing to the user (Nielson, 1993:26). Usability is a general concept that cannot be measured; however, it is directly related to usability criteria that *can* be measured (Nielsen and Levy, 1994:67). The emphasis of usability engineering is knowing exactly what criteria can and should be used to evaluate a product's usability (Dix and others, 1998:199). The discipline of usability engineering also provides a structured methodology for achieving usability in user interface design by evaluating these usability metrics through the development lifecycle of the information system (Mayhew, 1999:2; Whiteside and others, 1988:791).

Usability metrics fall into two broad categories: subjective user preference measures, which assess the degree of the *user satisfaction* of the target technology; and

objective performance measures, which assess *user capabilities* of utilizing the target technology (Nielsen and Levy, 1994:67). Although many questionnaires have been developed to assess a user's subjective satisfaction with a system, few have focused exclusively on user interface evaluations (Chin and others, 1988:213). The Questionnaire for User Interaction Satisfaction (QUIS), a commonly used and validated instrument for measuring user satisfaction (Chin and others, 1988), has proven useful in demonstrating the benefits of usability engineering (Schneiderman, 1998:134). Conversely, objective performance measures focus on the quantifiable determinants of usability. These determinants are measured as the effectiveness and efficiency of specified users which perform specific tasks in a given environment (Bevan and others, 1991; Dillon and Morris, 1998; Nielsen and Levy, 1994:67; Shackel, 1991:24; Smith, 1996:70; Whiteside and others, 1988:792.). Combining these subjective and objective usability metrics enables the usability engineers to easily establish usability goals for the user interface. For example: "Users should be able to perform a specified task with new tool after W minutes training, with X% effectiveness, at least Y% efficiency, and Z % greater satisfaction than the old interface where W < infinity, and 0 < [X,Y,Z] < 100" (Dillon and Morris, 1998).

The discipline of usability engineering provides demonstrable results to the question, "*can* users use the system?" However, with all the advantages usability brings to application design and development, it does not answer the question, "*will* users use a system?" For example, the system may be evaluated favorably on every performance measure, but the system may ultimately not be used because of user dissatisfaction with the user interface or some other aspect of the system (Chin and others, 1988:213).

Therefore, some have stated that usability is a necessary--although not a sufficient

determinant of actual use (Dillon and Morris, 1998).

The Management Information System (MIS) Approach

A popular model for predicting use of a computer technology was first introduced

by Davis, Bagozzi, and Warshaw (1989). This model (Figure 1), known as the

Technology Acceptance Model (TAM), is based on the tenets that *perceived usefulness*

(U) and *perceived ease of use* (EOU) are the fundamental determinants of user acceptance

behavior (Davis and others, 1989:985). U is defined as the extent that users believe the

new technology will improve their job performance (Davis, 1989:320). EOU is defined

as the degree to which a person believes that using a particular system would be free of

effort (Davis, 1989:320). User acceptance behavior, depicted in Figure 1 as *behavioral*

intention to use (BI), is operationalized in TAM as system usage. The
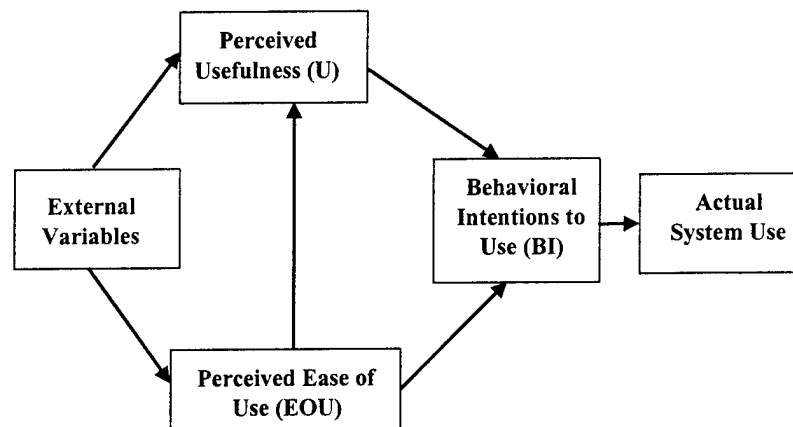
Figure 1. Technology Acceptance Model (TAM) (Davis, 1993)

external variables, which affect U and EOU in TAM, bridge a user's internal beliefs, attitudes and intentions (Davis and others, 1989:988). Also, usability determinants such as a well-designed graphical user interface, type of manual user input devices, system training, documentation, and user support may mediate the effects of EOU to greater extent than U (Davis and others, 1989:987-988; Venkatesh and Davis, 1996:473).

Previous research has successfully validated TAM as a viable model for predicting user acceptance of new and different information technologies over a wide range of organizational settings and technologies (Adams, Nelson and Todd, 1992; Chin and Todd, 1995; Davis and Venkatesh, 1995; Doll and others, 1998; Morris and Dillon, 1997; Szajna, 1994). TAM offers researchers and practitioners a relatively simple and cost effective way to predict the ultimate measure of system success, whether or not that system is actually used (Morris and Dillon, 1997:59).

TAM is a viable method of predicting actual usage of post production software applications. However, it is not entirely clear if TAM would prove useful for evaluating prototypes in the design phase of a software development effort (Morris and Dillon, 1998:64). Additionally, TAM is based solely on user perceptions and these perceptions over time can change considerably based on a long-term performance of the application and the dynamics of the work environment (Dillon, 1987).

Problem Statement

The usability engineering approach and the MIS approach offer valuable insights to information technology managers for predicting the success of implementing new information systems into an organization. However, no one approach captures the entire

7

dynamics involved in developing or implementing the best information system for an organization. TAM has reliably and repeatedly demonstrated a R^2, variance explained, ranging from 50% to 74% for a variety of office automation applications (Dillon and Morris, 1998); yet, TAM does not give system developers insights on how to build in acceptance features into new information technologies. Conversely, usability engineering does provide an effective methodology of developing useful user interfaces; however, usability does not guarantee user acceptance nor is it plainly understood how usability contributes to user acceptance.

It appears that a unified model that supports both the process of early design and clarifies the relationship between usability and acceptability would have a significant impact on reducing the risk of implementing new information systems (Dillon and Morris, 1998). The main risk, the rejection of a new information technology by users, can be mediated by incorporating acceptance features into the development phase of a new system. Additionally, incorporating acceptance features into the development phase could also potentially reduce the overall system maintenance costs by increasing the quality of the initial version of the system (Nielsen, 1993:7-8; Mantei and Teorey, 1988:428; Mayhew, 1999:449-450).

User acceptance research is predicated on a concept that *new technologies* which provide utility to an organization are evaluated for viability. Speech recognition software (SRS) is one new technology that has received notoriety as a viable tool for increasing productivity in an organization (Alwang, 1998:191; Toft, 1999). SRS enables computer systems to interpret and execute voice commands, and transcribe dictation from users into applications such as word processing programs, e-mail applications, and spreadsheets.

Additionally, SRS packages, which are now coming bundled in many office software suites (Morris, 1998:116), may soon find their way to desktop computers in Air Force offices. Therefore, SRS was primarily selected as the target technology because it meets the user acceptance research criteria due to its novelty and practical application to office automation software. Also, results from this research will indicate the viability of employing SRS in an Air Force office environment.

Research Questions

This study focuses on validating a new model for predicting actual system use as proposed by Dillon and Morris (1998). The new user acceptance model, entitled the Power, Performance, Perception (P^3) model, incorporates the subjective metrics and objective metrics of the usability engineering and MIS approaches previously outlined into an integrated model. The first factor, *power*, addresses the capabilities within the target software--typically operationalized as the number of features contained in application that contribute to utility. The second factor, *perception*, incorporates the subjective metrics of usability engineering and TAM. The third factor, *performance*, incorporates the objective metrics of usability engineering. Importantly, for the purposes of this research, if the P^3 model indicates that integrating these three factors yields a significantly higher R^2 than TAM, then the new model has the potential of decreasing the risk of implementing new information systems into organizations while simultaneously providing insights for application developers to design user acceptance attributes into new software applications. Therefore, this research addresses the following questions:

1. What is the significance of integrating usability engineering metrics and technology acceptance determinants into the proposed P^3 model on predicting user acceptance of an information technology as compared to TAM's ability to predict user acceptance?

2. If the P^3 model demonstrates as an enhanced predictor of information technology use, what usability factors significantly contribute to information technology acceptance?

3. What is the viability of incorporating speech recognition technology into an Air Force office environment?

This thesis reports on an experimental study using Dragon NaturallySpeaking SRS as the target technology to validate the P^3 model. A demographically diverse group of subjects was solicited to perform timed tasks utilizing SRS. These subjects were then administered a survey asking them about their perceptions of the software. The data gleaned from the timed tasks were used to obtain objective metrics while the data from the user survey were used to obtain the subjective metrics. Chapter II presents a review of literature relevant to the research; Chapter III describes the methodology used to collect research data; Chapter IV details the statistical analysis of the data, and Chapter V provides a discussion of the findings and conclusions.

II. Literature Review

Introduction

This chapter presents a comprehensive review of previous works in the two research areas of user acceptance and usability engineering of computer technology. The validated determinants of these two areas of concentration, user acceptance and usability engineering, form the constructs of the P^3, model which is the focus of this research. While there are conceptual overlaps in the literature sources of the two disciplines, the majority of the usability engineering literature is gleaned from HCI sources and user acceptance literature from MIS sources.

No one knows the exact amount of software applications, designed at a great cost of time and money, that are abandoned or expensively modified because the intended users do not use these applications after implementation (Markus, 1983:68). However, based on the proliferation of software in the workplace and the ever-increasing version numbers of the software applications, one can intuitively deduce that user expectations of software performance and capability are difficult to satisfy. Therefore, when software applications are implemented to increase the effectiveness and efficiency of users, information technology managers must consider many factors that influence adoption, also referred to as *user acceptance behavior*, of new technologies (Myers, 1997:21).

Evolution of User Acceptance Modeling

The importance and complexity of the adoption of new technologies was first popularized by the book <u>Diffusion of Innovations</u> (Rogers, 1983). In this book, Rogers

(1983:15-16) summarizes the key influence of user acceptance behavior as relative advantage, complexity, compatibility, trialability, and observability. The first key influence, relative advantage, is best described as the extent to which a potential adopter views the innovation as offering advantage over previous ways of performing the same task (Ararwal and Prasad, 1997:562). Complexity, the second key influence, represents the degree that an innovation is perceived to be difficult to understand, learn, or operate (Rogers, 1983:15). The third influence of acceptance behavior, compatibility, is determined by a degree that the innovation fits with a potential adopter's existing values, previous experience, and current needs (Rogers, 1983:16). Trialability, the fourth influence, is determined by the extent that potential adopters perceive that they have an opportunity to experiment with the innovation prior to committing to its usage (Ararwal and Prasad, 1997:562). And finally, observability, is the degree to which the results of an innovation are visible to others (Rogers, 1983:16). The information systems research within the context of innovation characteristics and adoption, as detailed by Rogers, views an individual's perceptions about these characteristics as important influences on user acceptance behaviors (Ararwal and Prasad, 1997:562; Moore and Benbasat, 1991:195; Taylor and Todd, 1995: 145).

An individual's perceptions or beliefs are also the basis of a widely studied model from social psychology entitled the Theory Reason Action (TRA), which was first proposed by Fishbein and Ajzen (1975). TRA is an especially well researched intention model that has demonstrated success in predicting and explaining behavior across a wide variety of domains (Davis and others, 1989:983). Some of the research domains utilizing TRA include education (Becker and Gibson, 1998; Koslowsky, 1993), consumer

behavior (Chang and Burns, 1996; Davis and Warshaw, 1991; Lee and Green, 1991),

alcohol dependence (Ulrich, 1990), criminal behavior (Welsh and Gordon, 1991), and

psychiatry (Currie and Aubry, 1995). In general, TRA is designed to explain virtually

any human behavior across a wide variety of situations (Ajzen and Fishbein, 1980:4).

TRA stipulates that a person's performance of a specified behavior is determined by his

or her behavioral intention (BI) to perform the behavior, and BI is jointly determined by

the person's attitude (A) and subjective norms (SN) concerning the behavior in question

(Davis and others, 1989:983) (Figure 2).
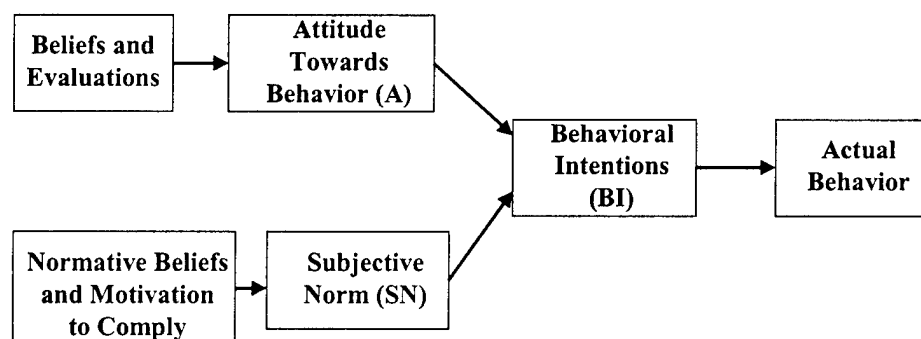
Figure 2. Theory of Reasoned Action (TRA) (Fishbein and Ajzen, 1980)

As depicted in Figure 2, A is formed by past experiences as a result of realized

consequences of past actual behaviors. For example, the resultant effect of drinking a cup

of coffee at 10:00 PM (actual behavior) was that it kept you awake until 3:00 AM the

following morning (consequence). Therefore, from a positive A perspective, if you

needed to stay awake until 3:00 AM, there is high probability that drinking a cup of

coffee at 10:00 PM would assist in accomplishing that objective. Thus, based on A, you

might have an increased intention to drink coffee. Looking at A from a negative

consequence perspective, if you needed to wake up at 6:00 AM, drinking a cup of coffee

at 10:00 PM the preceding night would result in an undesirable outcome of only three hours of sleep. In this case, your intentions of drinking coffee would be lower. This example illustrates how positive and negative attitudes are formed and then used as an antecedent to a behavioral intention.

The other antecedent to BI is subjective norm (SN). SN is defined as "the person's perception that most people who are important to him think he should or should not perform the behavior in question" (Fishbein and Ajzen, 1975:302). SN is a summation of two factors: a person's normative beliefs which are the perceived expectations of specific referent individuals or groups, and the motivation to comply with these expectations (Fishbein and Ajzen, 1975:302). In summary, TRA captures the internal psychological variables to which numerous external variables studied in MIS research achieve their influence on user acceptance (Davis and others, 1989:894). Studies involving TRA within the information systems domain include Davis and others (1989), Hartwick and Barki (1994), and Koslowsky and Hoffman (1990). Hartwick and Barki (1994) studied the effects of user participation on the system design process as it applies to actual system use. Unfortunately, the results of the Hartwick and Barki study did not provide specific design attributes that can be applied to system design.

An extension of TRA, the Theory of Planned Behavior (TPB) (Ajzen 1985, 1991) accounts for conditions where individuals do not have complete control over their behavior (Taylor and Todd, 1995:149). Similar to TRA, TPB states that actual behavior is a direct function of behavioral intention (BI) (Ajzen, 1991:181). However, the TPB model also includes an additional determinant, perceived behavioral control (PBC), which explains variance in BI and actual behavior above that provided by A and SN

(Ajzen, 1991:183) (Figure 3). PCB refers to one's perception of the ease or difficulty of performing the behavior of interest (Ajzen, 1991:183). For example, an individual may perceive he or she lacks the skill to use information technology and that skill level is important in determining the behavior of interest, usage (Taylor and Todd, 1995:150). A summary of 11 studies of TPB depicted in Ajzen (1991:187) indicates that the majority of results demonstrated that
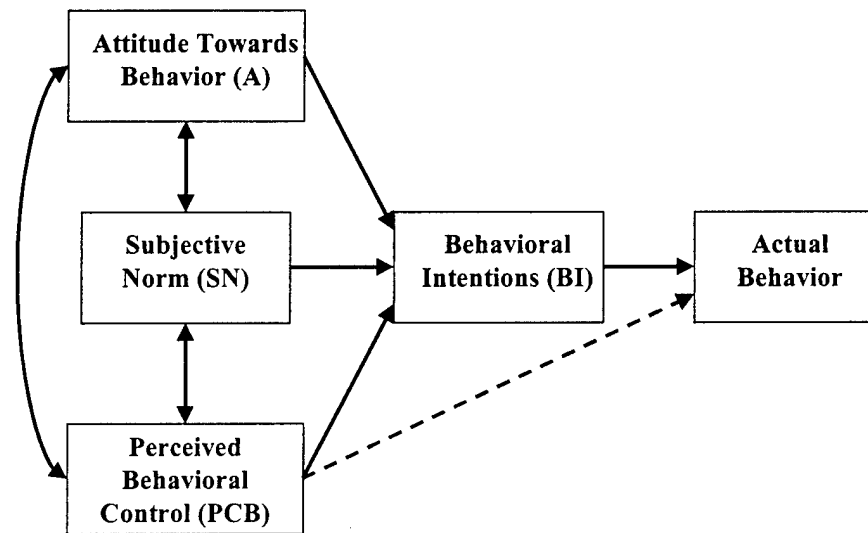
Figure 3. Theory of Planned Behavior (TPB) (Ajzen, 1991)

BI was a stronger predictor to actual behavior than PBC; however, PBC did significantly contribute to BI in various studies over SN and A. Therefore, the TPB regression model is represented as: BI = A + SN + PBC. Studies relevant to information systems research which incorporate TPB include Mathieson (1991), and Taylor and Todd (1995).

Of these three user acceptance research areas, innovation adoption, TRA, and TPB, TRA has immerged as a prominent model that has served as a basis for expanding user acceptance research. Specifically, a modified TRA model defined in the Davis and others (1989) study, resulted in a concise, complete, reliable and valid model to predict

user acceptance. This model, entitled the technology acceptance model (TAM), has repeatedly shown viability in predicting user acceptance of new and different information technologies over a wide range of organizational settings (Adams, Nelson and Todd, 1992; Chin and Todd, 1995; Davis and Venkatesh, 1995; Doll and others, 1998; Morris and Dillon, 1997; Szajna, 1994).

The Technology Acceptance Model

The *Technology Acceptance Model* (TAM) offers researchers and practitioners a relatively simple and cost effective way to predict the ultimate measure of system success, whether or not that system is actually used (Morris and Dillon, 1997:59). TAM is based on the tenets that *perceived usefulness* (U) and *perceived ease of use* (EOU) are fundamental determinants of user acceptance behavior (Davis and others, 1989:985). The effect on actual system use of these two fundamental determinants, U and EOU, are mediated through BI (Figure 4). Additionally, BI has been shown as an accurate predictor of actual use (Taylor and Todd, 1995:146).
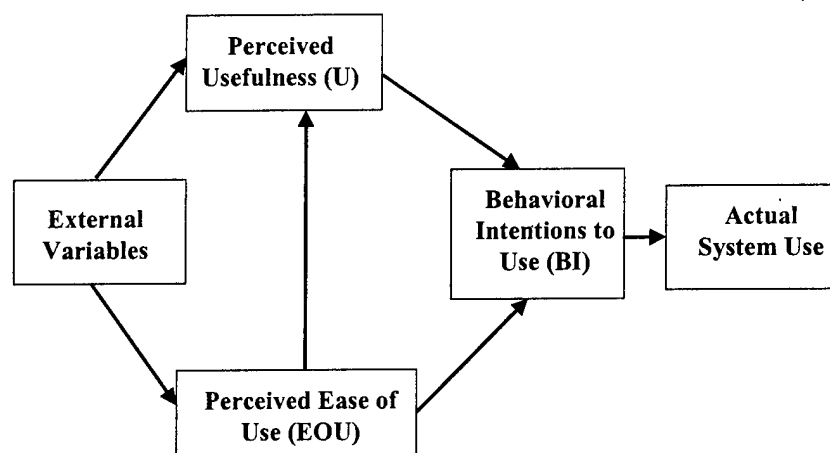
Figure 4. Technology Acceptance Model (TAM) (Davis, 1993)

16

TAM contends that people are inclined to use or not use the technology to the extent they believe it will improve job performance; this concept defines U (Davis, 1989:320). Additionally, even if potential users believe that a given application is useful, they may, at the same time, believe that the system is too hard to use and that the performance benefits of usage are outweighed by the effort of using the application (Davis, 1989:320). This perception leads to the second determinant of user acceptance, EOU, which is defined as the degree to which a person believes that using a particular system would be free of effort (Davis, 1989:320). Thus, EOU directly influences U (Davis, 1989:320). TAM's U-BI relationship illustrates that people form intentions towards using computer systems based largely on a cognitive appraisal of how it will improve their performance (Davis and others, 1989:986). Furthermore, the external variables represented in TAM which affect U and EOU bridge a users internal beliefs, attitudes and intentions (Davis and others, 1989:987-988). In addition, some determinants of usability such as a well-designed graphical user interface, manual user input devices such as a mouse or touch screen, training, documentation, learnability and user support may affect EOU to greater extent than U (Davis and others, 1989:987; Venkatesh and Davis, 1996:473).

Although TAM has been shown as an effective model for predicting user acceptance before a new technology is deployed in an organization, it does not provide guidance to system developers for designing user acceptance into an application (Mathieson, 1991:175). User acceptance research shows that BI is strongly determined by U and comparatively, EOU has a weaker, though still significant, effect on BI but

strongly influences U (Adams and others, 1992:229; Davis, 1993:476; Davis and others, 1989:996; Taylor and Todd, 1995:167). The empirical evidence also indicates that increasing the EOU of the system will increase U and the increased U will translate into an increased BI resulting in a larger margin of user acceptance; thus, EOU has a significant direct effect as well as an important indirect influence on BI through U. However, research also indicates that the influences of EOU on U diminish over time as users become proficient with the target system (Chau, 1996:197; Davis and others, 1998:996). Therefore, the literature suggests that EOU determinants will have the greatest contribution to user acceptance in the early stages of system deployment, when users have limited experience with a target system. For example, this concept of EOU and user acceptance as it applies to users that have limited experience with an application is specifically significant to users of the World Wide Web (WWW). As users visit unfamiliar WWW sites, they can quickly become discouraged if the site is not easy to use and quickly hyperlink to a similar WWW site regardless of the initial site's usefulness. Thus, the concept of EOU is the first hurdle of design and usage that system developers must address to gain user acceptance.

Evolution of Usability Engineering

Usability engineering is derived from the discipline of human computer interaction (HCI). HCI is a research area first defined as its own scientific discipline in the book The Psychology of Human-Computer Interaction, (Card and others, 1983), which the authors state that "certain central aspects of computers are as much a function of the nature of human beings as of the natures of computers themselves." Furthermore,

the authors state that the key to understanding the dynamics and complexities of the interaction between humans and computers is based on studying the methods that humans communicate with computers:

> The user and the computer engage in a communicative dialog whose purpose is the accomplishment of some task. It can be termed a dialog because both the computer and a user have access to a stream of symbols flowing back and forth to accomplish the communication; each can interpret, query, and correct the communication at various points in the process. All the mechanisms used in this dialog constitute the interface: the physical devices, such as keyboards and displays, as well as the computer's programs for controlling the interaction. (Card and others, 1983:4)

One of the primary goals of HCI is to enhance the quality of the interaction between people and computers to make that technology easier for people to learn and use (Baecker and others, 1995:1). This primary goal is the impetus of an applied form of HCI referred to as *user centered system design* (UCSD), which places the emphasis of computer systems design on the people using the systems rather than the technology (Norman and Draper, 1986:2). UCSD, also referred as *user centered design* (UCD), applies knowledge about human processes, human capabilities and limitations, and machine capabilities and limitations to the design process of computer systems to reduce the complexity of system operation and maximize a user's effectiveness with the system (Baecker and others, 1995:1; Gardner, 1991:134; Landauer, 1997:274).

The first step of UCSD is understanding who will be using the system. This step is commonly referred to as *user analysis* (Bannon, 1986:25-26; Gardner, 1991:136-138; Schneiderman, 1998:67-70). User analysis starts with compiling user profiles that consist of important attributes which may include age, gender, physical abilities, education, cultural-ethnic background, training, motivation, goals, and personality (Schneiderman,

1998:67-68). The next step of user analysis is understanding *how* users will utilize the system based on their skill levels. For example, first-time, novice, or infrequent users will have different behavioral characteristics and goals than expert users who frequently use the system (Schneiderman, 1998:68-69).

Once a thorough understanding of the people who will be using the system is attained, the next step is to perform a *task analysis*. The task analysis process is a collection of formal and informal techniques which defines: 1) the specific tasks the system is expected to perform to support an organization's processes, 2) how the system will perform these tasks, and 3) what effect the current and planned technologies will have on the tasks (Anderson and Olson, 1987:543; Landauer, 1997:276-280; Lewis and Rieman, 1993:123-124; Schneiderman, 1998:70-71). After the tasks that the system will perform are identified, the next step is to map these tasks to the type of users that will be performing the task on the system (Anderson and Olson, 1987:543; Landauer, 1997:279-280; Lewis and Rieman, 1993:124; Schneiderman, 1998:70-71). Understanding the user-to-task relationship is critical to the utility of the system, and the degree of utility is based on the quality of the user analysis and task analysis (Bonnon, 1986:26). The key to a quality user analysis and task analysis is interacting with the people who will be using the system, understanding how tasks are currently accomplished, and deducing how the new system can improve the efficiency and effectiveness of the people who will ultimately use the new system.

The final step in UCSD is the *iterative design process*. System developers apply the information gleaned from the user and task analysis to develop the system. However, it is highly unlikely that the first design of the system will adequately fulfill user

20

requirements or even perform properly (Lewis and Rieman, 1993:127). In the iterative design process, a prototype system is developed and actual end users who will ultimately use the system test the prototype. Redesign and future iterations of the prototype system are based on user perceptions and the successes and failures of the users performing specified tasks (Anderson and Olson, 1987:545; Lewis and Rieman, 1993:125-126; Landauer, 1997:281-283; Schneiderman, 1998:10). This iterative process normally continues until either an acceptable level of user and system performance is attained or the deadline for developing a system is reached (Anderson and Olson, 1987:547). However, missing from the UCSD iterative design process are clearly defined objectives that are the determinants of a usable system. Without these clearly defined objectives, the iterative design process will most likely perpetuate beyond a reasonable time schedule resulting in a system that is delivered late or never delivered.

Usability Engineering

Similar to UCSD, usability engineering can be divided into three main processes: user analysis, task analysis, and iterative design. However, usability engineering differs from UCSD in that explicit user and system performance metrics are introduced into the iterative design process to maximize the effectiveness and efficiency of system use (Dix, 1998:199, Gould and Lewis, 1987:528; Whiteside and others, 1988:792). Additionally, usability engineering is intended to integrate the principles of usability throughout the lifecycle of the product, with significant activities occurring in the early stages of development, and not left as a last-minute fix of the user interface before product release (Nielsen, 1993:71).

The usability of a system is typically measured by having a number of end-users, who are representative of the users who will ultimately use the system, perform a pre-specified set of tasks (Gould and Lewis, 1987:530; Nielsen, 1993:27). However, soliciting end-users to thoroughly test all aspects of all the versions of an evolving design can prove difficult or expensive (Nielsen and Mack, 1994:2). Therefore, another technique to detect and resolve usability problems is by employing *usability inspection* methods. Usability inspection is a generic name for a set of methods based on having the expert evaluators inspect or examine usability aspects of a system as opposed to end-users (Nielsen and Mack, 1994:1). Typically, a usability inspection focuses on revealing usability problems in an *existing*, post production, system and then using these problems to make recommendations for improving the usability of the system (Nielsen and Mack, 1994:3). Although usability inspections are a viable method of detecting and eradicating usability problems, integrating end-user usability testing from the early stages of application design is a more common method of empirically measuring usability throughout the development lifecycle.

End-user empirical usability testing largely depends on the objectives of the tasks identified in the task analysis process. Each task can be evaluated to the degree that objectives are met in the context of one, some, or all of the following usability attributes: learnability, efficiency, memorability, error tolerance, and user satisfaction (Nielsen, 1993:26). Usability acceptance criteria, measurable goals for system performance that must be attained prior to system delivery (Schneiderman, 1998:135), are assigned to each usability attribute specifying a range of performance measures a user must meet when using the system (Nielsen, 1993:80; Whiteside and others, 1988:794-797). Acceptance

criteria, or usability goals, are normally established prior to the system design process to give developers concrete targets to assess their design ideas against as they generate and consider concepts for implementation (Mayhew, 1999:123).

The first attribute of usability, learnability, is measured in the time required for a user to learn how to accomplish a task or set of tasks on the system (Chapanis, 1991:363, Nielsen 1993:29). In some cases, learnability is the most fundamental usability attribute, since the first experience most people have with a new system will involve learning to use the system (Nielsen, 1993:27-28). Another method for measuring learnability consists of specifying that users must complete a set of tasks in a specified minimum time before being considered them as having *learned* the system (Nielsen, 1993:29).

Efficiency, the second usability attribute, is a performance metric reflecting the level of productivity once the user has learned to use the system (Nielsen, 1993:26). Efficiency is typically measured in the time it takes a user to perform a specific set of tasks (Chapanis, 1991:363; Nielsen, 1993:31). For example, during the iterative design process, user performance is repeatedly measured in terms of the number of seconds to accomplish a specific task. When performance has not increased for some iterations, the users is assumed to have reached the steady-state level of performance for the tasks specified (Nielsen and Philips, 1993:214-221).

The third usability attribute, memorability, focuses on *casual users* of the target system (Nielsen, 1993:31) and measures user retention of commands overtime (Schneiderman, 1998:135). "Casual users are people who are using the system intermittently rather than a fairly frequent use assumed for expert users" (Nielsen 1993:31). There are two methods for conducting memorability tests: user performance

tests and memory tests. The first method, user performance tests, involves casual users who have been away from the system for a specified amount of time and are then timed while performing a set of tasks (Nielsen, 1993:32). The second method, memory tests, requires the casual user to perform a specific set of tasks and then answer questions about the effects of various commands or to name the commands used to accomplish those tasks (Nielsen, 1993:32).

Ideally, a system should prevent users from performing erroneous actions and, should errors occur, give clear instructions to the user on how to restore the system to its previous state without corrupting any previous work (Schneiderman, 1998:75). This is the premise behind the fourth usability attribute, error tolerance. A system error is defined as any action that does not accomplish the desired goal of the user regardless of whether it is an operator error or a defect in the software code (Nielsen, 1993:32). Error tolerance is measured by the number of errors a user performs or encounters while attempting to accomplish a specific set of tasks or the time it takes a user to recover from errors while attempting to accomplish a specific set of tasks (Chapanis, 1991:363; Nielsen, 1993:32-33).

Closely related to efficiency and error tolerance, in terms of system performance, is the concept of system *effectiveness*. Effectiveness is defined as the extent that the intended the goals of using the overall system can be achieved (Smith, 1996:70). Effectiveness can be measured as the percentage of the specified range of tasks completed by the user utilizing the target system in terms of speed and errors (Shackle, 1991:25). For example, users X, Y and Z successfully completed an average of 75% of tasks 1 through 20 within 30 minutes while experiencing four errors. Usability acceptance

criteria for effectiveness may be specified that users must be able to complete 95 percent of the tasks within 25 minutes without experiencing any errors.

Although a system may be evaluated favorably on every usability attribute, users may still not use the system because they are dissatisfied with the system or dissatisfied with its user interface (Chin and others, 1988:213). This concept is the premise for the final usability attribute, subjective satisfaction. Subjective satisfaction is normally measured using a validated instrument that canvasses a user, after he or she has a chance to thoroughly interact with the target system, about the user's subjective impressions about specific aspects of the system (Nielsen, 1993:34; Schneiderman, 1998:133). A number of questionnaires can be used to measure user subjective satisfaction. For example, IBM's Post-Study System Usability Questionnaire focuses on system usefulness, information quality and interface quality (Lewis, 1995:60), while the Software Usability Measurement Inventory is designed to measure a user's perception of their effect, efficiency, and control utilizing the system (Kirakowski and Corbett, 1993:210). One of the most popular questionnaires used by many usability professionals is the Questionnaire for User Interface Satisfaction (QUIS), developed by Chin and others (1988) of the University of Maryland. QUIS contains five sets of subjective questions measuring a user's overall reaction to the software, user interface screen characteristics, user interface terminology and system information messages, learnability, and system capabilities (Chin and others, 1988:215).

A common misconception in the usability field is that merely following a standard set of guidelines will result in a usable system. Although compliance with standards or guidelines will normally contribute to usability, it will not guarantee the usability of a

product (Bevan and others, 1991; Landauer, 1997:294; Morris and Dillon, 1996:244). The problem with usability guidelines is that they are often contradictory, hard to follow, and sometimes so lengthy that system developers refuse to use, or improperly use, the guidelines (Landauer, 1997:294-295; Molich and Nielsen, 1990;338). If guidelines are used properly they can prove valuable in promoting consistency in the design among multiple developers (Schneiderman, 1998:79); however, guidelines should not be relied on as the only method for incorporating usability attributes into the system development process.

Linking Usability Engineering to User Acceptance

The field of usability engineering provides significant guidance in developing usable systems by integrating the concepts of UCSD with usability attributes in designing systems to meet specific usability goals and has successfully answered the question, "*can* users use the system?" However, developing a *usable* system does not guarantee that the system will be *accepted* by end-users. Conversely, TAM is a reliable method of predicting system usage prior to full-scale system implementation and answers the question "*will* users use a system?" Unfortunately, TAM fails to provide specific guidance to application developers on system design. This important distinction of *"can versus will* users use the system?" suggests that neither research discipline sufficiently captures the scope of system design and implementation.

Several studies have modified TAM, or modified other user acceptance models, by incorporating TAM's user acceptance determinants to more precisely predict system usage (Agarwal and Parsad, 1997; Chau, 1996; Igbaria and Guimaraes, 1995; Jackson

26

and others, 1997; Taylor and Todd, 1995). However, none of the studies clearly addresses the issue of how the determinants of user acceptance can be integrated into the system design process.

The P³ Model: Bridging Usability Engineering and User Acceptance

The P³ model (Figure 5) proposes a unified model of predicting system use that supports both the process of system development and clarifies the relationship between usability and acceptability (Dillon and Morris, 1998). The P³ model is composed of three main components which act as determinants to a user's behavioral intentions (BI) to use a system. Furthermore, the first determinant, *power*, is based on a system's functionality that directly support the tasks used in the evaluation. The two other determinants, *performance* and *perception*, are affected by a unique set of antecedents that are derived from the usability engineering and user acceptance paradigms.
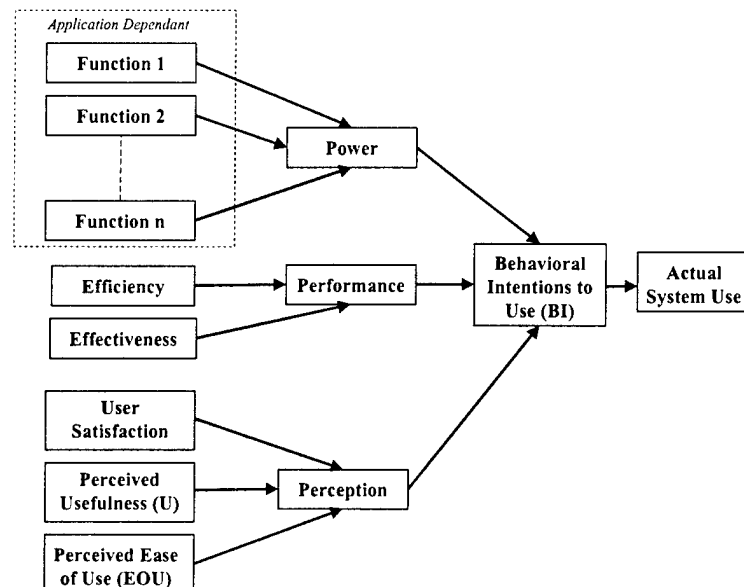
Figure 5. The Power, Performance, and Perception (P³) Model

Power, the first determinant of the P^3 model, is derived from the concept of a system's *utility*. Utility refers to the technical capability, or functionality, of an application that supports the accomplishment of a specific set of tasks that are salient to the user (Dillon and Morris, 1998; Nielsen, 1993:25). Additionally, the utility of an application is also dependent on the amount of different functions in an application that are available to the user the support the accomplishment of a specific set of tasks (Chapanis, 1991:363). This concept of enumerating the functions of an application has been operationalized as *function points* in the software engineering community. Additionally, function points are commonly used as a metric to estimate software size and complexity (Pressman, 1992:48). Although the total number function points in an application are not a measure of *power*, specific function points that directly affect the accomplishment of a specific task might possibly be highly correlated to *power*. Therefore, a system's power is unique to the application and is objectively measured as the number of functions utilized by the user to perform the target set of tasks (Dillon and Morris; 1998). For the purposes of this research, power is decomposed into the functionality required to perform basic word processing: text generation, text formatting, and text editing.

The second determinant of the P^3 model, performance, captures two objective metrics of usability engineering, efficiency and effectiveness, which best describe the operational relationship between the user and the target system. Additionally, efficiency and effectiveness are specified as the primary objective usability metrics in ISO 9241: Ergonomic Requirements for Office Work with Visual Display Terminals; Part 11- Guidance on Usability (Smith, 1996:69-70). Measures of effectiveness correspond to the

usability goals that reflect the level of accuracy and completeness achieved by the user on the target system (Bevan, 1997; Shackel, 1991:25; Smith, 1996:70). Efficiency, simply stated, is the resources expended to achieve the intended usability goals; in particular for this research, efficiency is measured by the amount of time it takes the user to complete a specific range of tasks in a given environment (Nielsen, 1993:31; Shackle, 1991:25; Smith, 1996: 70). For example, one measure of effectiveness can be the amount of errors in a document depending on the input mode, SRS or keyboard. If there are 10% fewer errors in the document produced with the SRS as compared to the document produced with only a keyboard, one can deduce, in terms of errors, SRS is 10% more effective. Furthermore, a measure of efficiency can be determined by comparing the time it takes someone to generate a text document using speech recognition software (SRS) and the time it takes to generate a similar document using only a keyboard. Comparing the number of words in each document to the time required to generate each document, one could conclude which method, SRS or keyboard, was more efficient for text generation.

The final determinant of the P^3 model, perception, combines the subjective measures of usability engineering and user acceptance to describe a user's attitude towards the system (Dillon and Morris, 1998). User acceptance research has shown that perceived usefulness and perceived ease-of-use directly influence user attitude (Davis and others, 1989:985). Additionally, usability engineering operationalized user attitude as the subjective measure of user satisfaction towards a system (Bevan, 1997; Chin and others, 1988:213; Nielsen, 1993:33-37; Smith, 1996:70). Understanding the interrelationship between these two methods of subjectively measuring user attitude and objectively

measuring performance leads to the identification of which usability attributes significantly contribute to user acceptance.

Target Technology: Speech Recognition Software (SRS)

The inherent value of user acceptance modeling, in particular TAM, is its ability to reliably predict system usage behavior of new technologies within an organization prior to any large capital investments of the new technologies under scrutiny. One new technology that has recently received much notoriety as a viable tool for increasing productivity in an organization is speech recognition software (SRS) (Alwang, 1998:191; Toft, 1999). SRS enables computer systems to interpret and execute voice commands, and to transcribe dictation from users into applications such as word processing programs, e-mail applications, and spreadsheets. The transcription of dictation into word processing type applications is known as speech-to-text processing, and controlling the operations of a computer system using speech commands is referred to as command and control (Markowitz, 1996:183). Although current SRS capabilities are far from achieving this seamless integration and functionality that one might see dramatized in the Star Trek series, SRS packages are appearing in office software suites (Morris, 1998:116) and may soon find their way to desktop computers in Air Force offices. Therefore, utilizing SRS for this research serves two purposes:

1. SRS meets the new technology criteria for user acceptance modeling due to its novelty, availability, and practical application to office automation software and

2. SRS can be examined for viability in an Air Force office environment.

The October 20, 1998 issue of PC Magazine evaluated the four leading

commercially available SRS packages: Dragon NaturallySpeaking 3.0, IBM ViaVoice 98

Executive, L&H Voice Xpress Plus 1.01, and Philips FreeSpeech98. PC Magazine

concluded that Dragon NaturallySpeaking was superior in accuracy and ease-of-use for

speech-to-text processing and IBM ViaVoice performed best for command and control

functions. Overall, the magazine's product evaluators recommend that Dragon

NaturallySpeaking is the best package due to its 91% average accuracy for speech-to-text

capabilities. Both Dragon NaturallySpeaking and IBM ViaVoice currently retail for

$150. For the purposes of this research, Dragon NaturallySpeaking was selected over

IBM ViaVoice because of the Dragon product's superior speech-to-text capabilities.

Shortly after Dragon shipped NaturallySpeaking, the first general-purpose

continuous-speech recognition application, in June 1997 (Poor, 1998:131), the software

was evaluated in by the Air Force Communications Agency Directorate of Technology

(AFCA/TC) (Dennis, 1998). The study consisted of five users who assessed the

transcription accuracy and command recognition of the software over a six week period

(Dennis, 1998:2). The results of the study concluded that the software was not mature

enough for full implementation throughout the Air Force because of the software's

cumbersome operating speed and low transcription accuracy rate of 75% (Dennis,

1998:7). The findings of the AFCA/TC are questionable due to the size of the subject

pool, n = 5. Additionally, the assessment methodology described in the report lacks the

controls of an experimental setting. However, considering this was the first version of

Dragon NaturallySpeaking and the insufficient processing power of the computers

running the software, some measure of validity can be derived from the study's conclusions.

The findings of a more recent and academically rigorous study, conducted by a human-computer interaction research group, of the usability of SRS were recently presented at the Association of Computing Machinery's Human Factors in Computer Systems Conference (Karat and others, 1999). This study involved 24 subjects who performed text composition and dictation tasks using both SRS and keyboard-mouse modes of text entry (Karat and others, 1999:570). The three commercially available large vocabulary, continuous-speech recognition systems were used in this study: IBM ViaVoice 98 Executive, Dragon NaturallySpeaking Preferred 2.0, and L&H Voice Xpress Plus (Karat and others, 1999:569). Subjects were divided into three groups and each one of the groups was assigned one of the three SRS packages mentioned above (Karat and others, 1999:570). The results of this study indicated that subjects took longer completing the tasks when using SRS as the primary input mode, found SRS more difficult to use, and believed they would be less or much less productive with SRS than with keyboard and mouse (Karat and others, 1999:574-575). Although this study provided details on both objective and subjective usability metrics, the study's overall statistical power was diluted due to having only eight subjects per treatment. Stronger statistical power can be achieved applying at least 30 subjects per treatment as prescribed by the Central Limit Theorem (McClave and others, 1998:254). Additionally, since the objective usability metrics were pooled across the three groups and then the results averaged, it was difficult to tell if one of the groups using a particular SRS package

weighted down the overall average score. In other words, it was not clear whether one, two, or all three SRS packages performed poorly.

It is obvious that SRS will greatly assist the physically impaired in exploiting the capabilities of PCs as well as the hunt-and-peck typists who languish over a computer keyboard every time they compose a document. However, it is yet to be accurately determined if SRS technology can significantly increase office productivity and more importantly, be accepted by users once deployed in an organization.

Understanding the Influence of Usability Engineering on User Acceptance

The evolution of usability engineering from the discipline of HCI has established the basis for the research that has answered the question, "*can* users use the system?" Conversely, from the discipline of MIS, user acceptance research has answered the question, "*will* users use the system?" However, neither discipline has successfully produced conclusive evidence that explains what characteristics facilitate system use as well as influence a user's acceptance of the system. This study examines the validity of a new model, P^3, that may link the constructs of usability engineering to user acceptance and thus provide system developers insights on incorporating usability features, that significantly contribute to user acceptance, into applications early in development. Additionally, SRS, selected as the target technology due to its novelty and practical application to office automation software, is utilized in an experimental setting to gather data to validate the P^3 model. Also, this research examines the viability of employing SRS in an Air Force office environment.

III. Method

Research Approach

This chapter describes the methodology used to acquire the data which were analyzed to address the research questions presented in Chapter I. The leading commercially available speech recognition software (SRS) package was the target technology used by human subjects in an experimental setting to validate the proposed P^3 model. The ensuing sections describe the subject pool, experimental design, and instrumentation and statistical analysis used in this study.

Subject Pool

The 30 subjects used in this experiment were active duty Air Force members and civil service employees working for the Air Force located a major Midwestern U.S. Air Force installation. To qualify as a subject for this experiment, each subject met the following prerequisites:

1. Work in an office environment on a daily basis and utilize Microsoft Word 97 in the office to compose written documents,

2. must not have any experience utilizing any type of SRS application, and

3. must be free of any sinus infection or nasal congestion as not to adversely affect voice quality.

All subjects who participated in this experiment did so voluntarily and did not receive any financial compensation for participation. Subjects were informed that the

individual results gleaned from their participation in the experiment would be kept confidential and result specifics would not be attributed to any individual participant. Additionally, subjects were also informed that the research would focus on the capabilities of the software and not the ability of the subject to show proficiency using the software. Last, all subjects were advised that they had to right to terminate their participation in the experiment at any time.

Experimental Design

The experiment consisted of two sessions, each lasting approximately 1.5 hours, with each session conducted on separate days. The experiment was divided into two sessions, as opposed to one three-hour session, based on results from pilot studies which indicated a threat to internal validity as subjects became fatigued after the second hour of experimentation. Therefore, subjects were required to complete both 1.5 hour sessions within a 48 hour period. If a subject could not complete the second session within the allotted timeframe, the subject was eliminated from the study. Each session consisted of one researcher leading the experiment with one subject performing the tasks. Observational tasks were individually timed for each subject; however, subjects were not given a time limit to complete the tasks. To ensure that each subject received identical guidance from the researcher, scripts were developed and used by the researcher which detailed what was communicated to the subject in each session of the experiment (Appendices 1, 2 and 3).

The first session, day one of the experiment, consisted of the subject training the SRS application to recognize and interpret the subjects voice and also train the subject to

use SRS. Training the SRS, also referred to as *enrollment,* is an automated feature inherent to all commercially available SRS applications. The enrollment feature guides the user through a dictation task from a predefined script for approximately 20 minutes and then automatically creates the user-specific speech files used by the software to interpret the user's voice. Once enrollment is complete, the computer system was re-started to clear the random access memory (RAM) buffers which were heavily taxed during the enrollment process. It was discovered in the pilot studies that re-starting the computer system after enrollment improved the performance of the SRS speech-to-text processing functionality. The next step following enrollment required subjects to review a multimedia computer-based training (CBT) module inherent to the SRS application. The CBT instructed the subject on the fundamentals of using SRS to perform speech-to-text entry and using voice commands to perform text editing and text formatting operations. Portions of the CBT, which included using voice commands to navigate within the document, advanced text formatting techniques, and advanced mis-recognition error correction techniques, were omitted because these features were not utilized in the experiment.

Following completion of the CBT, subjects were required to perform a simple speech-to-text task. The purpose of this task was to familiarize the subject with the SRS by using a simple text selection that did not require any text formatting such as indenting, bold, underline or italics. First, the researcher handed the subject a paper copy of a 141 word paragraph (Appendix 4) and instructed the subject to dictate the paragraph into the system. The system consisted of the Pentium II-350 MHz Dell desktop computer with 128 MB RAM running Microsoft Word 97 and Dragon NaturallySpeaking Professional

3.01, and a Labtec C-316 headset microphone. All paper copy handouts were similarly formatted to the following specifications: line spacing-1.5, font style-Times New Roman, font size-12 point, top and bottom margins-one inch, and sides margins -.8 inches. A blank document was pre-formatted on Microsoft Word by the researcher to the above specification prior to the subject beginning the task. Additionally, the automated grammar checker function was disabled but the automated spell checker was left enabled.

The next phase in this session involved the researcher training the subject to utilize voice commands for text editing and text formatting. The majority of these commands were presented in the CBT; however, the CBT only demonstrated these commands and did not permit the subject to utilize them. In this task the researcher handed the subject a paper copy of the table (Appendix 5) containing 20 voice commands which were used by the subject to format and edit the text in the 141 word paragraph previously dictated. The researcher gave a brief synopsis of each command individually and then instructed the subject to perform the command. Once a command was successfully completed by the subject, the researcher preceded to the next command on the table. This cycle continued until all 20 voice commands were accomplished by the subject and the subject felt comfortable using all the commands.

The final phase of session one consisted of the subject inputting a text passage similar in context and format, although shorter, to the text passage that was used on day two of the experiment. This text passage was dictated from a handout provided to the subject from the researcher (Appendix 6). The subject was instructed to input and format the text as it appeared in the handout; additionally, the subject was also instructed that in the event of a mis-recognition error, a subject could make only three attempts to use voice

commands to correct the error. Mis-recognition errors occur when the SRS system wrongfully interprets the utterances of the user and generates incorrect text. If, after three attempts the mis-recognition error was not corrected by the SRS, the subject was instructed to continue dictating and ignore the error. Furthermore, while the subject was dictating and formatting the passage, the researcher was permitted to advise the subject on techniques or voice commands to facilitate accomplishing the task. Also, during the previous phase or this final phase of session one, subjects were encouraged to ask questions to clarify any issues regarding SRS use. Once this task was completed, the subject was reminded of the day and time of session two, instructed not to discuss the experiment with other potential subjects, and excused.

The second session, day two of the experiment, required each subject to perform two timed tasks consisting of text generation and formatting. One task required the subject to utilize conventional input modes of keyboard and mouse to enter and format text from a handout provided by the researcher. The other task required the same subject to enter and format text from a different handout provided by the researcher utilizing SRS and limited mouse functions. Mouse functionality was restricted in the SRS task to document navigation and text selection. Additionally in the SRS task, subjects were given the option to select text by either using voice commands or the mouse.

The two tasks were constructed to isolate and control five variables of interest. These variables included a subject's experience with SRS, the word processing functionality of the Microsoft Word 97, the familiarity to the content of each handout by the subject, the word/character count of each handout, the formatting count of each handout, and a text input mode. The objective of the tasks was to manipulate the text

input mode variable, between conventional and SRS modes, while holding the other four variables constant. Therefore, the affects of a subject's productivity and perceptions could be accurately measured based on the input mode.

A subject's experience with SRS was controlled by first screening out subjects who had experience with SRS and then ensuring that all subjects had the same amount of SRS training. SRS training was kept constant by exposing subjects to the same amount and type of training on day one of the experiment.

Word processing functionality was held constant for each subject by identically configuring the default document format settings of font type and size, line spacing, and margins for each task. Documents included the blank document used by the subject to input and format text and the actual handouts provided to the subject from the researcher. Additionally, as stated previously, the automated grammar checker was disabled and a spell checker was enabled for each subject. Since each subject completed two tasks involving text entry and formatting with varying input modes between the tasks, a subject's familiarity with the content of the handout's text attained by inputting text in the first task could affect the performance of text entry in the second task. Therefore, to control for text familiarity, the content of the text was changed for each task to ensure that the subject was equally unfamiliar with the text in each task. Although the content of the text was dissimilar in each handout, the amount of text and the amount of formatting was held approximately constant (Table 1). Handout 1 (Appendix 7) was always administered in the first task and handout 2 (Appendix 8) was always administered in the second task. Lastly, manipulating the input mode variable posed a problem of contending with the order of the task performed, conventional text input mode first or SRS text input

39

mode first. It was not known whether the sequence of performing the task would affect the subject's productivity or perceptions between tasks. Therefore, to minimize order effect, the sequence of modes was alternated for each subject. For example, subject one would perform the conventional text input mode task first and then would perform the SRS text input mode task. Subject two would then perform the SRS mode task first followed by the conventional mode task. Furthermore, subjects were randomly assigned to perform one of the two orders of inputting text.

Table 1. Handout Statistics

	Handout 1	Handout 2
Words	606	570
Characters	3156	3190
Paragraphs	12	13
Lines	54	59
Bold Functions	8	11
Underline Functions	8	7
Italic Functions	2	0
Hyphen functions	4	8
Hard Return Functions (equivalent to "enter-key")	16	16
Indent Functions (equivalent to " tab-key")	4	4

Each task in the second session was divided into two parts: part one required the subject to input text using one of the two input modes and part two consisted of proofreading the entered text. Each part of the two tasks accomplished by the subject was individually timed and recorded by the researcher. Additionally, prior to starting the SRS task, the subject was given the opportunity to practice text inputting and formatting with

the SRS. For this SRS practice task, the subject was instructed to dictate and format a text passage from a handout provided by the researcher (Appendix 9). Subjects were also reminded, in the event of a mis-recognition error, to give the SRS three chances to correct the error and then continue dictating. While the subject was dictating and formatting the text, the researcher could advise the subject on SRS text input techniques or voice commands. Additionally, subjects were encouraged to ask questions to clarify any issues regarding SRS use. However, during the actual SRS timed task, researchers were not permitted to coach the subject and subjects were not permitted to ask questions.

Once both tasks were completed, the subject was asked to complete a questionnaire (Appendix 10) regarding the subject's perceptions of using SRS. Details of the questionnaire are explicated in the instrumentation section of this chapter. After the questionnaire was completed, the researcher briefed the subject on the purpose experiment and the task performance times recorded. Last, subjects were instructed not to discuss the experiment with other potential subjects and then excused.

Instrumentation and Statistical Analysis

Proper evaluation, which is consistent with the literature, of all three components the P^3 model, *power*, *performance*, and *perception*, consisted of decomposing each individual component into its determinants and then measuring the individual determinant's effect on the dependant variable, behavioral intention (BI). The first P^3 determinant, *power*, was decomposed into the functionality required to perform basic word processing: text generation, text formatting, and text editing. For this research, power was held constant across each task and was not directly measured since the word

processing functionality of the application, Microsoft Word 97, did not vary between tasks. Therefore, the effects of *performance* and *perception* on BI were isolated and measured using instrumentation and methodologies from the areas of usability engineering and TAM research.

Objective usability metrics, which are composed of the quantifiable measurements of usability goals reflecting efficiency and effectiveness, support the *performance* construct of the P^3 model. Therefore, the first step in evaluating performance was establishing usability goals for efficiency and effectiveness. The following are the usability goals for this research:

1. Subjects, utilizing SRS with limited mouse functionality, must *accurately* perform the task of inputting and formatting text from a handout at an equal to or higher rate than performing a similar task with manual input modes of keyboard and full mouse functionality.

2. Subjects, utilizing SRS with limited mouse functionality, must achieve a text inputting and formatting error rate equal to or lower than an error rate utilizing manual input modes of keyboard and mouse.

3. Subjects, utilizing SRS with limited mouse functionality, must achieve an equal to or higher accurate words per minute (AWPM) rate than performing a similar task with manual input modes of keyboard and full mouse functionality.

The first usability goal, an effectiveness metric, was determined in terms of *accurate words*. The count of accurate words was derived by the number of *correctly spelled* and *correctly intended* words in the final document produced by the subject at the conclusion of each task. The concept of *correctly spelled words* is self-explanatory;

however, the concept of *correctly intended words* is a new concept that must be accounted for when determining the accuracy of the speech recognition system. For example, SRS rarely generates misspelled words or produces typographical errors which are common to manual keyboard entry methods. SRS technology is prone to misrecognition errors where a user may utter a word like "speech" but the SRS may generate the text "beach." Although the word "beach" is spelled correctly, it is not the word the user intended. Therefore, to derive the number of *accurate words* generated by the SRS, words in the final document must meet the criteria of being correctly spelled and correctly intended.

The second usability goal, also an effectiveness metric, was derived from the number of errors in the final document. Errors were separated into six different categories. The first error category (ec_1) consists of word spacing errors. Word spacing errors were identified by the excess or deficiency of spaces between words and punctuation marks. For example, the beginning of every sentence following a period should start with two spaces from the period of the previous sentence to the first word in the following sentence. One spacing error was attributed to the error category if exactly two spaces were not present after the period of a preceding sentence and before the first word of the following sentence.

Line spacing errors comprise the second error category (ec_2). Line spacing errors consist of the incorrect number of line spaces between sentences. For example, although line spacing was defaulted to automatically input 1.5 lines between sentences for soft and hard returns, subjects were required to input consecutive hard returns on occasion

43

throughout the two tasks. If the subject failed to input the appropriate amount of hard returns, one error was accumulated for each hard return not entered.

Capitalization errors, the third error category ($ec3$), were measured as the amount of words that were mis-capitalized. For example, if the first word of the sentence was not capitalized or a word that should be all lowercase letters is capitalized, then one capitalization error is accumulated for each occurrence.

Text formatting errors ($ec4$) consist of text characters that were improperly formatted or missing a required format. Specifically, text formatting includes characters that are bolded, italicized, or underlined. In this error category, one error was counted for each word that was improperly formatted or missing a required format.

Inaccurate words are the next category on errors (ec_5). This category of errors is measured by the amount of inaccurate words in a final document produced by a subject performing one of the tasks. Inaccurate words were categorized as words that were either misspelled, spelled correctly but were not the intended word, or were superfluous.

Omitted word errors (ec_6) are measured as the total number of words that were omitted in a final document produced by a subject performing one of the tasks. The last category of errors (ec_7) accounts for the total number of punctuation errors, which include incorrect and omitted punctuation.

A subject's overall effectiveness for the corresponding usability goals was determined by calculating the difference between the objective measures of the SRS task and manual input mode task (Table 2). The formula to derive the overall effectiveness rating for usability goal one (UG-1) and usability goal two (UG-2) was designed to yield a positive value if the goal was met or a negative value if the goal not met. The objective

measures for UG-1 was derived by taking the total number of *accurate words* (AW) in the resultant document of each task for each subject and dividing it by the *total number of words in the handout* (TWHO) used by the subject to generate the resultant document. The objective measures for UG-2 were calculated by the summation of the total number of errors in each error category (ec_i) for each task accomplished by each subject.

Table 2. Effectiveness Metric Chart

		(1) SRS Task	(2) Manual Task	(3) Overall Effectiveness
UG-1	% Accurate Words Per Document	$\dfrac{AW}{TWHO}$	$\dfrac{AW}{TWHO}$	(1) − (2)
UG-2	# Errors Per Document	$\sum_{i=7} ec_i$	$\sum_{i=7} ec_i$	(2) − (1)

The overall efficiency measure of each subject for usability goal three (UG-3) was calculated by deriving the difference between the objective measures of the SRS task and manual input mode task (Table 3). Similar to UG-1 and UG-2, the formula to derive the overall efficiency rating for usability goal three (UG-3) was designed to yield a positive value if the goal was met or a negative value if the goal not met. The objective measures for UG-3 were derived by taking the total number of accurate words (AW) in the resultant document of each task for each subject and dividing it by the time, in minutes, it took the subject to generate and proof read the resultant document.

Table 3. Efficiency Metric Chart

		(1) SRS Task	(2) Manual Task	(3) Overall Efficiency
UG-3	Accurate Words Per Minute (AWPM)	$\dfrac{AW}{Minutes}$	$\dfrac{AW}{Minutes}$	(1) – (2)

Measurement of the *perception* construct of the P³ model consisted of subjective measures TAM and usability engineering research. Each subject was administered a questionnaire at the conclusion of session two of the experiment (Appendix 10). The questionnaire was composed of validated instrument scale items which subjectively measure user perceptions of SRS. A total of eight scale items were adapted from TAM research (Davis, 1989): four scale items which measured *perceived ease-of-use* (EOU) and four items which measured *perceived usefulness* (U). All eight of these TAM research scale items each used a 7-point Likert scale, where a level seven response, indicating a high level of its concurrence with a scale item, to a level one response, indicating a severely weak concurrence with the scale item. The overall measures of EOU and U were derived by calculating the average Likert scale responses for the four scale items associated with each determinant. Additionally, the Cronbach Alpha for the scale items were evaluated to measure item reliability and accuracy.

The post experiment questionnaire also contained 13 user satisfaction scale items derived from usability engineering research (Chin and others, 1988; Lewis, 1995). Eight of the user satisfaction scale items were extracted from the Questionnaire for User Interface Satisfaction (QUIS) (Chin and others, 1988) and five were extracted from the IBM Computer Usability Questionnaire (Lewis, 1995). Although there were 27 scale

items available in the QUIS instrument and 19 scale items available in the IBM

instrument, most items were tailored to assess a graphical user interface and could not

directly translate to voice user interface measurement. The QUIS scale items were scored

by the subject using a 10 level scale ranging from zero as the most negative score to nine

as the most positive score. The five scale items from the IBM questionnaire were scored

on a 7-point Likert scale. Similar to the TAM research scale items, user satisfaction scale

items were averaged to derive an overall score, and Cronbach Alpha's were also evaluated

to test for scale item reliability. Additionally, two scale items were used to measure

users' behavioral intentions (BI) to use the SRS system if implemented in their

organizations (Davis and others, 1989).

Finally, measurements gleaned form the post-experiment questionnaire and the

objective usability metrics were used in a hierarchical multiple regression (HMR)

analysis. HMR analysis develops a self-weighting estimation equation which predicts

values for a dependent variable from the values for several independent variables (Cooper

and Schindler, 1998:562). The standardized coefficient estimations, also known as beta

weights (β), of the independent variables, indicate the degree of affect to the dependent

variable. Also derived from HMR, and of particular relevance to the research, is the

variance explained measurement, R^2. As stated in Chapter I, the viability level of the P^3

model, as compared to TAM, was determined by examining the resultant R^2.

Furthermore, descriptive statistics were scrutinized to gain additional insights on usability

issues and trends related to SRS use.

IV. Results

Results

This chapter details the results of the experimental study described in the previous chapter. The demographic information of subjects was gathered from a pre-experiment questionnaire (Appendix 11), subjective measurements were attained from post-experiment questionnaires (Appendix 10) and objective measures for each subject were recorded by the researcher at the completion of each task during the experiment. The data from the post-experiment questionnaires and the objective measures were complied and analyzed through regression testing and factor analysis to examine the validity of the P^3 model. The findings of the analyses are encapsulated in Tables 1 through 9.

Demographic Information

The content of Table 4 displays the numerical demographics of the subjects who participated in the experiment. The composition of the subject pool is representative of a typical Air Force office environment at the managerial staff or junior executive level.

Table 4. Subject Demographics

	Subjects	Male	Female	Enlisted	Officer	Civilian
Totals	30	23	7	6	20	4
Mean Age	31.50	30.57	34.57	30.50	29.55	42.75
Minimum Age	22	22	24	22	25	24
Maximum Age	67	54	67	38	38	67
	9.19	6.77	15.02	6.89	3.82	21.19

All subjects reported that they had no experience using the target software

package and two subjects reported that they were briefly exposed to a speech recognition

software (SRS) package other than the target package, but did not consider themselves

proficient using any SRS package. Last, Table 5 depicts the self-reported average time

estimates of the hours spent using MS Word in a typical work week.

Table 5. Average Time Spent per Week Using MS Word

Hours	None	1-5	5-10	10-20	>20
Subjects	0	8	13	8	1

Task Performance Results

Table 6 presents the overall results of the objective time measurements for the two

main tasks performed by each subject. The data for the keyboard tasks are displayed in

the column labeled "Key" and the data for the SRS tasks are in the column labeled SRS.

The column labeled "Δ" contains data indicating the differences between the keyboard

and speech tasks. Values used to determine the calculations for the "Δ" column were

derived by subtracting each subject's SRS task time from the keyboard task time.

Table 6. Subject/Task Performance Time Summary

	Key	SRS	Δ
Mean	0:26:48	0:27:38	0:06:19
Max	0:39:59	0:43:26	0:19:11
Min	0:16:35	0:14:39	0:00:12
σ	0:05:43	0:07:42	0:04:38

The content of Tables 7 and 8 display the objective performance metrics based on the accuracy of each of the two main tasks. Errors for the keyboard task are displayed in the column labeled "K_Errors" and errors for the SRS task are displayed in the "S_Errors" column. Columns labeled "K_APWM" and "S_AWPM" display the accurate words per minute (AWPM) data for the keyboard and SRS task, respectively. AWPM are calculated by dividing the total number of accurate words generated in each task by the time elapsed to complete that task. Columns labeled "K_APWD" and "S_AWPD" display the accurate words per document (AWPD) data for the keyboard and SRS task, respectively. AWPD percentage data is displays in columns "K_APWD%" and "S_AWPD%". Percentage data are calculated by dividing the AWPD by the total amount of words that should have been generated.

Table 7. Subject/Task Performance Accuracy Summary 1

	K_Errors	S_Errors	K_AWPM	S_AWPM
Mean	8.00	21.37	24.12	22.43
Max	35.00	69.00	36.37	38.43
Min	1.00	6.00	14.08	13.24
σ	7.98	14.01	5.57	6.80

Table 8. Subject/Task Performance Accuracy Summary 2

	K_AWPD	S_AWPD	K_AWPD%	S_AWPD%
Mean	586.57	572.00	100%	97%
Max	606.00	603.00	100%	100%
Min	563.00	533.00	99%	89%
	18.20	22.45	0.37%	2.17%

Reliability and Factor Analysis

The data displayed in Table 9 reflect the scale item and variable validity for the post-experiment questionnaire. Variables measured in the questionnaire are the determinates for the *perception* component of the P^3 model. These *perception* variables include perceived Ease of Use (EOU), perceived Usefulness (U), Attitude (A), Questionnaire on User Interface Satisfaction (QUIS), and IBM's Post-Study System Usability Questionnaire (IBM). The primary dependent variable, Behavioral Intention (BI), is utilized by the P^3 model as the key predictor of actual usage.

The data in the column labeled "Loadings" contains the loadings of the factor analysis for each scale item. Loadings account for the variance between scale items for each variable. Columns labeled α_O, α_L, and α_U display the Cronbach Alpha reliability values for each variable. Alpha values evaluate the consistency of responses between scale items for each variable. The alpha values in the table below are divided into three columns: the overall values (α_O), values for the lower fiftieth percentile range (α_L), and values for the upper fiftieth percentile range (α_O). The upper are lower fiftieth percentile ranges were based the subjects response to the scale items for BI. The response range for BI is based on a seven point Likert scale where "1" represents the most favorable response (lower number limit) to the scale item, and "7" represents the most unfavorable response (upper number limit). A crossloading discrepancy on question (Q1) resulted in the elimination of that scale item for the EOU measurement. Only the overall result data for QUIS and IBM were analyzed.

The results from the reliability and factor analysis indicate a strong inter-item correlations among scale items. Generally, researchers prefer Alpha values of .60 or

greater for behavioral studies (Morris and Dillon, 1997:61). The alpha values for this study all exceed .75, indicating a strong reliability of the scale items. TAM scale items loaded strongly (>.86) on a single factor for each of the scale items (BI, EOU, U, and A), as expected, with low cross-loadings. Loadings for the QUIS and IBM scale items also demonstrated high correlation as they ranged from .763 to .876 and .664 to .906, respectively.

Table 9. Reliabilities and Factor Analysis

Variable	Question	Loadings	α_T	α_L	α_U
BI			0.987	0.932	0.978
	Q22	0.994			
	Q24	0.994			
EOU			0.833	0.757	0.754
	Q3	0.863			
	Q6	0.862			
	Q9	0.922			
U			0.973	0.950	0.964
	Q2	0.963			
	Q7	0.954			
	Q10	0.975			
	Q13	0.955			
A			0.936	0.870	0.952
	Q5	0.964			
	Q8	0.940			
	Q12	0.960			
	Q16	0.910			
QUIS			0.911	-	-
	Q51	0.876			
	Q52	0.844			
	Q53	0.782			
	Q54	0.846			
	Q55	0.848			
	Q57	0.791			
	Q58	0.763			
IBM			0.835	-	-
	Q59	0.777			
	Q60	0.748			
	Q61	0.744			
	Q62	0.664			
	Q63	0.906			

P³ Model Linear Regression Analysis

The primary focus of this study was to examine the user acceptance predicting ability of the P³ model by integrating usability engineering metrics with validated TAM determinates. The first component of the P³ model, *power*, was held constant for this study and was not analyzed. The second component, *performance*, is composed of efficiency and effectiveness usability metrics. Efficiency was measured by the variable total time per document (TIME). Effectiveness was measured by the variables errors per document (ERR) and accurate words per document (AWPD). The effects of these *performance* variables on BI, as well as the other *perception* variables, are displayed in Table 10. This table is divided into three major partitions: "SRS", "Δ", and "|Δ|". The "SRS" partition contains data exclusively reflecting the subjects' SRS performance metrics. The "Δ" partition contains data reflecting the difference between keyboard performance metrics and SRS performance metrics. The "|Δ|" partition contains data reflecting the absolute value of the difference between keyboard performance metrics and SRS performance metrics. Each partition in Table 10 contains values for variance explained (R^2), beta coefficients (b), t-value (t), and significance (p). For clarity, horizontal lines separate each set of variables analyzed. Dependent variables are annotated in the column labeled "DV" and independent variables are in the column labeled "IV."

The results in Table 10 indicate that the *performance* variables are inconsequential determinates of BI and the other P³ *perception* variables. The low R^2 values (< .20) of the *performance* variables indicate a weak effect on BI and the high p-values (> .05) of the independent variables show minor significance levels. Additionally,

based on R^2 and p-values, the results also demonstrate that the *performance* variables

failed to even have a minor effect on the other *perception* variables.

Table 10. Performance Metrics Analysis

DV	IV	SRS				Δ				|Δ|			
		R^2	b	t	p	R^2	b	t	p	R^2	b	t	p
BI		0.095				0.061				0.072			
	ERR		-0.177	-0.69	0.499		-0.381	-1.15	0.260		0.485	1.36	0.184
	AWPD		-0.356	-1.48	0.152		0.242	0.76	0.455		-0.303	-0.93	0.360
	TIME		0.163	0.80	0.433		-0.011	-0.05	0.959		-0.079	-0.35	0.732
EOU		0.096				0.122				0.125			
	ERR		-0.088	-0.34	0.737		-0.550	-1.72	0.098		0.613	1.77	0.088
	AWPD		-0.309	-1.29	0.210		0.303	0.98	0.336		-0.279	-0.88	0.385
	TIME		0.170	0.83	0.416		0.047	0.24	0.814		-0.219	-0.99	0.334
U		0.021				0.055				0.074			
	ERR		-0.100	-0.37	0.714		-0.374	-1.13	0.271		0.474	1.33	0.194
	AWPD		-0.184	-0.73	0.470		0.329	1.03	0.315		-0.413	-1.27	0.214
	TIME		0.036	0.17	0.867		0.000	0.00	0.999		-0.016	-0.07	0.944
QUIS		0.207				0.135				0.166			
	ERR		0.057	0.23	0.823		0.417	1.28	0.211		-0.524	-1.52	0.142
	AWPD		0.308	1.31	0.202		-0.133	-0.42	0.675		0.269	0.86	0.400
	TIME		-0.363	-1.83	0.079		0.099	0.50	0.625		-0.103	-0.47	0.646
IBM		0.081				0.142				0.186			
	ERR		0.038	0.14	0.890		-0.536	-1.66	0.110		0.579	1.70	0.102
	AWPD		-0.205	-0.81	0.424		0.242	0.78	0.446		-0.337	-1.09	0.288
	TIME		0.141	0.66	0.513		-0.001	-0.01	0.996		0.112	0.51	0.613

The third and final component of the P^3 model, *perception*, is composed of two

user satisfaction variables, QUIS and IBM, and two TAM variables, EOU and U. Table

11 displays how these perception variables affect BI and how the TAM variables affect

the individual user satisfaction variables.

The results of the analysis indicate that the TAM variables are strong determinates

of BI and the user satisfaction variables, when analyzed concurrently with the TAM

variables, show less the moderate effects on the dependant variable. The values of the

beta coefficients of the TAM variables indicate that EOU has a slightly

Table 11. Perception Metrics Analysis

DV	IV	R^2	b	t	p
BI		0.884			
	EOU		0.698	4.72	0.000
	U		0.428	4.25	0.000
	QUIS		0.146	1.15	0.261
BI		0.884			
	EOU		0.664	5.27	0.000
	U		0.469	4.44	0.000
	IBM		-0.146	-1.19	0.244
BI		0.494			
	QUIS		-0.703	-5.14	0.000
BI		0.533			
	IBM		0.730	5.56	0.000
BI		0.560			
	QUIS		-0.302	-1.251	.222
	IBM		0.476	1.970	.060
QUIS		0.709			
	EOU		-0.856	-5.48	0.000
	U		0.019	0.12	0.902
IBM		0.692			
	EOU		0.621	3.87	0.001
	U		0.262	1.63	0.115

larger effect on BI than U. However, when the user satisfaction variables were

individually examined for effects on BI, the QUIS and IBM variables demonstrated as a

strong determinant to BI. The beta coefficient for QUIS is negative because the scale is

inverted, 0 is the least favorable rating and 9 is the most favorable rating. Additionally,

EOU showed significant deterministic effects on both QUIS and IBM.

SRS User Acceptance Linear Regression Analysis

The secondary focus of this study was to examine the viability of incorporating SRS technology into an Air Force office environment. Determining viability can be divided into two steps. The first step involves assessing the predicting ability of the TAM variables. Once TAM is validated, the second step consists of evaluating the mean responses of the TAM variables, specifically BI, and inferring the variance explained (R^2) from the first step.

The data contained in Table 12 verifies the user acceptance predicting ability of the TAM variables. Table 12 is vertically divided into three major partitions: "Overall," "Lower," and "Upper." The "Overall" partition contains the data reflecting the entire subject pool. The data were divided in half, based on the value of BI, to gain a further understanding of the relationship the independent variables and the dependent variable. Results gleaned from this division of data will indicate the degree of effect each independent variable had for the subjects who where more likely to use SRS and for the subjects who were least likely to use SRS. Thus, values displayed in the "Lower" partition reflect the responses to BI of the lower fiftieth percentile of the subject pool representing the most favorable responses to the dependant variable. The values displayed in the "Upper" partition reflect responses of the upper fiftieth percentile of the subject pool representing the least favorable responses to the dependant variable.

The "Overall" results indicate that the TAM variables explain 88% of the variance. The variance explained in this study is consistent with Davis (1998; $R^2 = .74$), Doll and others (1998; $R^2 = .70$), and Jackson and others (1997; $R^2 = .67$). Additionally, the results support the removal of the attitude construct variable (A) as a determinate to

BI and mediator of EOU and U because of the negligible impact to R^2 and high p-value. The omission of A is also consistent with the findings of Davis (1993) and Venkatesh and Davis (1996).

Table 12. TAM Variables Analysis

DV	IV	Overall				Lower				Upper			
		R^2	b	t	p	R^2	b	t	p	R^2	b	t	p
BI		0.876				0.524				0.846			
	EOU		0.565	5.79	0.000		0.333	1.63	0.130		0.500	3.24	0.007
	U		0.443	4.53	0.000		0.570	2.78	0.017		0.503	3.26	0.007
BI		0.883				0.527				0.856			
	EOU		0.437	3.06	0.005		0.299	1.20	0.255		0.359	1.62	0.133
	U		0.360	3.06	0.005		0.533	2.09	0.060		0.407	2.15	0.054
	A		0.213	1.23	0.231		0.078	0.26	0.797		0.241	0.89	0.390
BI		0.800				0.298				0.762			
	A		0.895	10.59	0.000		0.546	2.35	0.035		0.873	6.45	0.000
A		0.851				0.507				0.819			
	EOU		0.604	5.64	0.000		0.436	2.09	0.058		0.585	3.50	0.004
	U		0.386	3.61	0.001		0.471	2.26	0.043		0.399	2.39	0.034
U		0.520				0.054				0.461			
	EOU		0.721	5.51	0.000		0.233	0.86	0.403		0.679	3.34	0.005

The "Lower" results indicate that U is a major determinant of BI for the subjects who rated BI most favorably. Additionally, EOU did not demonstrate as a significant determinant for U for the lower fiftieth percentile group. Conversely, both TAM variables demonstrated as strong determinants in the upper fiftieth percentile group. EOU and U demonstrated having an approximately equal effect on BI, b = .500 and b = .503, respectively.

The subjects' mean rating of BI was a score of 3, indicating that they "slightly agree" they would use SRS technology in their work environment should that technology be made available to them. The data examined in this chapter indicates that a subject's perception of the ease of use (EOU) and usefulness (U) of the target technology adequately explains a subject's behavioral intention (BI) to use that technology. Thus, the validity of the scale items used to measure the TAM variables is effectively supported by the results of the reliability and factor analysis (Table 9). Therefore, since BI has been shown as an accurate predictor of actual use (Taylor and Todd, 1995:146) and based on that EOU and U explain 88% of the variance BI, it can be inferred that SRS technology would be utilized in a typical Air Force office. However, the performance metrics summarized in Tables 6, 7, and 8 indicate that in the short term, average users would not experience an increase in text production; in fact, users would be likely to experience a slight decrease in productivity.

Summary of Results

The low R^2 values and the high p-values of the *performance* variables, ERR, AWPD and TIME, on the dependent variable, BI, indicate that the objective usability metrics were ineffective predictors of user acceptance. The *performance* variable results also failed to demonstrate any significant effects on the *perception* variables, EOU, U, QUIS, and IBM.

Of the *perception* variables, the TAM variables EOU and U preformed strongly as determinates of BI. Additionally, the beta coefficients indicate that EOU has a slightly more significant effect on BI than U. Conversely, the other *perception* variables that

measure user satisfaction, QUIS and IBM, demonstrated as s strong determinants to BI when analyzed individually for effects on the dependent variable. Also, QUIS and IBM were significantly effected by EOU.

Last, the strong performance of TAM suggests that the results from the questionnaire can be used to predict user acceptance of SRS technology in an Air Force office environment. Furthermore, subjects' responses indicate they "slightly agree" they would use SRS technology, although the objective measures show that they would be slightly less productive with its use.

V. Discussion

General Discussion

The main focus of this study was to validate a new model, P^3, for predicting user acceptance of new technology. Additionally, if the P^3 model was validated, its determinants could also potentially identify objective performance metrics and subjective satisfaction metrics that significantly contribute to user acceptance. Thus, by identifying the salient objective and subjective determinants that influence users to accept new technology, it can guide system developers to develop new applications that have an increased probability of being accepted.

One of the most influential MIS research models that has repeatedly performed well in predicting user acceptance, TAM, exclusively utilizes subjective metrics that support the *perceived ease of use* (EOU) and *perceived usefulness* (U) constructs. These two constructs are very effective for predicting user acceptance and have answered the questions, "*will* users use the system?" However, user perceptions such as EOU and U provide minimal specific design guidance for developing new systems that will be accepted by users. Conversely, usability engineering, a HCI based discipline, effectively utilizes objective and subjective metrics to assist system developers in creating *usable* applications and has focused on the question, "*can* user use the system?" Still, developing applications with a usability engineering approach does not guarantee that users will use the application once deployed. Therefore, bridging the gap between developing usable systems and systems that will be used, would significantly contribute the body of knowledge of HCI and MIS researchers and practitioners.

This study reports on the results of the empirical data gleaned from experimental research that explored the validity of a new user acceptance model entitled the Power, Performance, Perception (P^3) model. The P^3 model integrates objective and subjective metrics from both the usability engineering and MIS disciplines (Table 13). The objective metrics measure a users' performance utilizing the target technology and the perceptions metrics measure user satisfaction, perceived ease of use, and perceived

Table 13. Variable to Metric Relationships

	Performance/Objective Metrics		Perception/ Subjective Metrics	
	Efficiency	Effectiveness	TAM	Satisfaction
Usability	TIME	ERR, APWD	-	QUIS, IBM
MIS	-	-	EOU, U	-

usefulness of the target technology. The P^3 model is built on a solid theoretical foundation and strong intuitive appeal. Thus, it would seem that this model has the potential to reveal the yet unidentified connection between usability and user acceptance. Determining the user acceptance predicting viability of the P^3 model motivated the researcher to pursue the answer to the following question:

What is the significance of integrating usability engineering metrics and technology acceptance determinants into the proposed P^3 model on predicting user acceptance of an information technology as compared to TAM's ability to predict user acceptance?

Examining the results of the *performance* component of the P^3 model indicates that the *effectiveness* and *efficiency* variables do not significantly contribute to a user

behavioral intention to use (BI) the target technology (Figure 6). BI acted as the

dependant variable (DV) for the P^3 model and served as the measurement of, "will users

use the system?" Several factors may have contributed to the weak effects of the

performance variables. First, the power of the effectiveness variables, error rates (ERR)

and accurate words per document (AWPD), and the efficiency variable, time to complete

task (TIME), may have been affected by the experimental design that required subjects to

perform both manual keyboard entry and SRS tasks in session two of the experiment.

Subjects may have formed inaccurate perceptions due to fatigue induced by performing

both tasks in one 1.5 hour session. Perhaps if subjects only preformed the SRS task in

session two, there would have been a higher correlation of the *effectiveness* and *efficiency*

variables with the BI measurements. Also, subjects may have developed inaccurate

perceptions of their performance with the target technology based on high expectations of

the capability of the SRS.

b = -0.177, p = 0.499

ERR

b = -0.356, p = 0.142   R^2 = 0.095

AWPD ——————→ BI
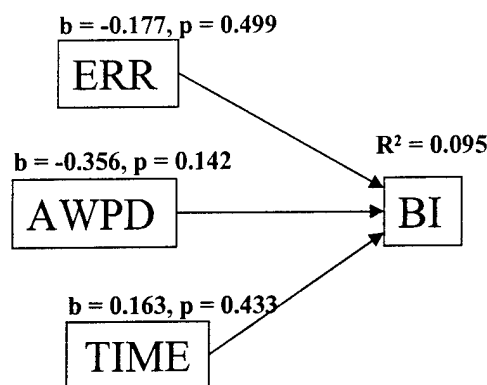
b = 0.163, p = 0.433

TIME

Figure 6. Effects of Performance Metrics on BI

Second, the performance results also suggest ERR may explain some variance in

the subjective metrics of EOU, IBM's Post-Study System Usability Questionnaire (IBM),

and the Questionnaire on User Interface Satisfaction (QUIS). The data in Table 11 indicate that ERR may have some notable effects on the perception variables when the SRS and manual performance metrics were combined. The values displayed in the three "b" columns of Table 11 represent the deterministic power of ERR on the dependent variables (DV). Examining the significant increases in the deterministic power of ERR and significance levels, "p", after the SRS and manual performance metrics were combined, shows that the subjects' performance on the manual keyboard tasks may have affected the subjective metrics of EOU, IBM and QUIS. Although none of the significance levels attained a value that suggested ERR was a salient determinant to the relevant DV ($p < 0.05$), these increases in significance indicate that the manual keyboard tasks may have affected the subjects' response to the perception variables.

As expected, the negative values of ERR's beta coefficients (b) for EOU and IBM in the "SRS and Manual (Δ)" partition of Table 14 indicate that as the error rate decreases, user satisfaction and a user's perceived ease of use of the target technology increases. Additionally, the scale value measurements for QUIS are opposite in value to the other perception metrics; therefore, an increase in the value for QUIS represents a lower level of user satisfaction. Thus, the results indicate that the as error rate increases, QUIS will reflect a decrease in user satisfaction.

Table 14. Effects of Error Rate on Perception Variables

DV	IV	SRS Only			SRS and Manual (Δ)			SRS and Manual (|Δ|)		
		R2	b	p	R2	b	p	R2	b	p
EOU		0.096			0.122			0.125		
	ERR		-0.088	0.737		-0.550	0.098		0.613	0.088
QUIS		0.207			0.135			0.166		
	ERR		0.057	0.823		0.417	0.211		-0.524	0.142
IBM		0.081			0.142			0.186		
	ERR		0.038	0.890		-0.536	0.110		0.579	0.102

The results of the *perception* component of the P^3 model indicate that the TAM variables, EOU and U, and the user satisfaction variables, QUIS and IBM, contribute to the variance explained of a user's behavioral intention to use (BI) the target technology. However, the relationships between the independent and dependant variables are not necessarily in alignment with the construct of the P^3 model.

The TAM variables were separately analyzed with each of the user satisfaction variables, QUIS (Figure 7A) and IBM (Figure 7B), to test for effects on BI. The results indicate that EOU and U explain most of the variance ($R^2 = .884$) and that QUIS and IBM do not significantly affect BI. The strong performance of EOU and U was expected because these variables have consistently demonstrated a high degree deterministic effect on BI in previous TAM research. It was unclear whether QUIS and IBM would significantly affect BI since QUIS and IBM where developed as *stand-alone* instruments to measure satisfaction of user interfaces as opposed to the TAM variables which were developed as determinates to BI. Based on results, QUIS and IBM failed to demonstrate a significant effect on BI beyond that already accounted for by the TAM variables.
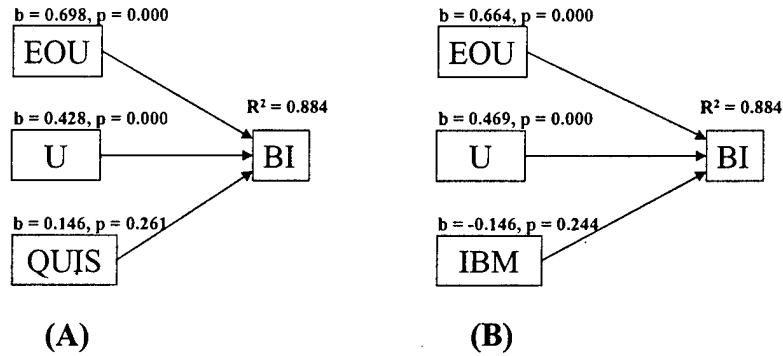
Figure 7. Effects of Perception Metrics on BI

However, in examining whether the user satisfaction variables might mediate the influences of the perception variables (U and EOU), the user satisfaction variables did perform strongly as determinants to BI when modeled as the sole determinant to the independent variable (Figure 8A and 8B). Of the two satisfaction variables, IBM obtained a slightly higher beta coefficient and variance explained than QUIS. When
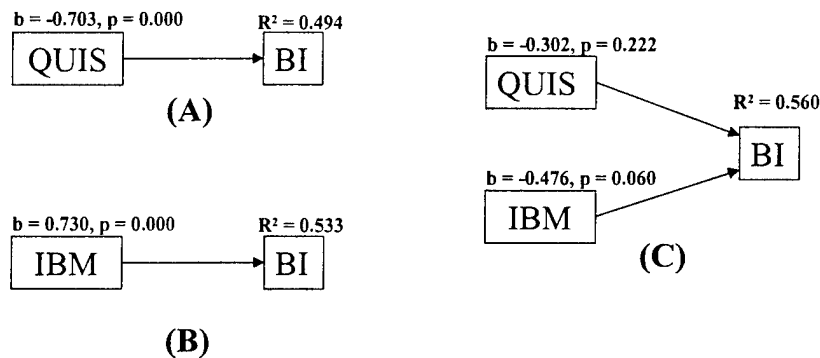

Figure 8. Effects of User Satisfaction Variables on BI

modeling both QUIS and IBM simultaneously, the QUIS variable significance diminishes as a determinant to BI (Figure 8C). Considering the higher performance of the IBM variable when evaluated individually and concurrently with QUIS, it can be concluded

65

that IBM could be used as viable single variable determinate to BI as IBM can explain 73% of the variance with only five scale items.

Upon further evaluation of the user satisfaction variables, EOU demonstrated that it significantly acts as a determinant to both QUIS and IBM. However, U failed to perform significantly on QUIS (Figure 9A) and U demonstrated weak significance on IBM (Figure 9B). EOU's strong deterministic effects on both user satisfaction variables suggest that the easier a user interface is to use, the more a user will be satisfied with the target technology. However, previous TAM research indicates that as users become more proficient utilizing the target technology, the effects of EOU will diminish and U will become the dominate determinant. Therefore, the model represented in Figure 8B may potentially be used to reliably predict user satisfaction throughout the lifecycle of the target technology.
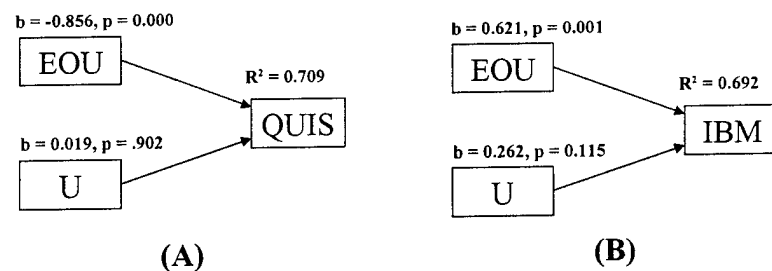

Figure 9. Effects of TAM Variables on Satisfaction Variables

Understanding the relationship between the usability metrics and user acceptance, the secondary focus of the this study, motivated the researcher to investigate the answer to this question:

If the P^3 model demonstrates as an enhanced predictor of information technology use, what usability factors significantly contribute to information technology acceptance?

Although this study failed to demonstrate an increase in the user acceptance predicting power of the P^3 model as compared to TAM, this study revealed some salient observations regarding the subjective satisfaction metrics. The significance of identifying salient subjective usability metrics that influence users to accept new technology can guide system developers to develop new applications that have an increased probability of being accepted. This study did identify that QUIS and IBM, when individually modeled as the sole determinate to BI (Figure 8A and 8B), explained 49% and 53% of the variance, respectively. This finding suggests that the extent to which a user is satisfied with the application's interface can be directly correlated with a user's behavioral intention to use the system. Furthermore, the data indicate that a user's *perceived ease of use* (EOU) of the user interface significantly affects user satisfaction. However, the EOU construct does not provide specific guidance on developing systems that will be accepted by user. Therefore, the data suggest that if system developers use the results of the scale items of the QUIS or IBM variables as metrics to improve user interfaces in the early stages of development, those improvements would result in a proportional increase in the probability that users would use system.

Of the two user satisfaction variables, the IBM variable may potentially prove as a more influential determinant to BI in a longitudinal setting, as QUIS does not show any indication of mediating the effects U. TAM literature suggests as users become proficient utilizing a system, the effects of EOU diminish and U becomes the dominant determinant.

Since IBM mediates the effects of EOU and U, the IBM variable emerges as the salient

user satisfaction variable effecting BI (Figure 10B). Additionally, IBM mediates the

effects of EOU and U much like the attitude construct (A) of TAM. A measures a user's

attitude formed by using the system. However, TAM's attitude construct does not

provide any insights for system development unlike the IBM variable that does provide

usability metrics for user satisfaction. Therefore, IBM could replace A creating a new

model called the Usability-Acceptance Model (Figure 10) that would provide insights to

applications developers that would contribute to user acceptance early in the development

process.
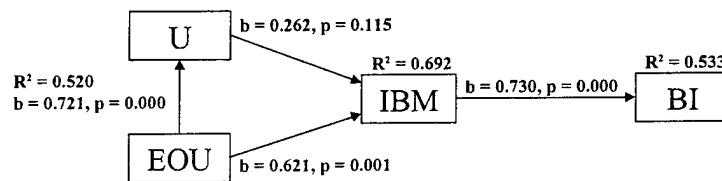
Figure 10. Usability-Acceptance Model

Examining the practicality of the target technology used in this study, speech

recognition software (SRS), motivated the researcher to explore the answer of the

following question:

What is the viability of incorporating speech recognition technology into an Air

Force office environment?

The results of this study suggest if SRS technology was deployed in an Air Force

office environment, users would utilize SRS even though the performance data indicate a

slight decrease in text processing productivity. Prior to the experiment, subjects

exclusively used keyboard and mouse to produce text-processed documents.

Surprisingly, after only 1.5 hours of training, subjects were able to produce text-processed documents with SRS with a mean accuracy rate of 97%. Additionally, subjects on average approximately took 6 minutes longer to produce the document with SRS. Although this study showed that productivity was slightly reduced with the use of SRS, productivity would likely increase as users became more experienced and comfortable using SRS.

Supplemental questions on the post-experiment questionnaire were used to gather some general user perceptions on the subjects' preferences on which modality, manual or SRS, they preferred for accomplishing specific text processing functions. A 7-point Likert scale was used in which a response of 1 indicated a strong preference for the speech recognition modality to 7, indicating a strong dislike of the speech recognition modality, for the specific function. The results in Table 15 suggest that subjects were indifferent to the modality for entering text and slightly preferred the traditional manual methods for text formatting and text selection.

Table 15. Text Processing Modality Preferences

	Text Entry	Text Formatting	Text Correction	Text Selection
Mean	4	5	5	5
σ	1.66	1.63	1.68	1.43

Subjects indicated that they perceived that it was slightly easier to correct typing errors than it was to correct mis-recognition errors. Four of the subjects wrote in the comments section of the questionnaire that mis-recognition errors were more difficult to

identify and correct than spelling errors. A possible explanation for this difficulty is that

unlike MS Word which automatically underlines misspelled words in red,

NaturallySpeaking SRS does not provide any overt visual cue of mis-recognition errors.

One subject wrote, "I'm usually aware of spelling errors as I make them and can correct

them right away. I wasn't aware of some of the mis-recognition errors, so proofreading

was more difficult." Alluding to the lack user feedback of mis-recognition errors, another

subject wrote, "Also because of the mis-recognition problem, you have to more

thoroughly proof read the document, missing the trust element." This "trust element" is

salient observation with respects to manual text entry. Users are confident that when they

press the A-key on the computer keyboard the letter A will appear on the screen. This

last comment infers that when the word "beach" is uttered, one is not entirely sure

whether "beach" or a similar sounding word like "peach" will be generated by the SRS.

This lack of trust of the software would normally require users to have a higher state of

vigilance of the output of SRS than they would have of outputs of manual modalities.

However, trust or confidence with SRS would likely increase as users become more

proficient with SRS.

Implications

Although the results of this study did not support the validity of the P^3 model, an

alternate model for predicting user acceptance, the Usability-Acceptance Model (Figure

10), did emerge as having both theoretical and practical implications for system

development. From a theoretical stand point, the Usability-Acceptance Model

demonstrated that the usability metric of user satisfaction significantly affects the MIS

construct of a user's behavioral intention to use a new technology. This finding might be the first indication of a theoretical model that may begin to explain the relationship between, "can users user the system?" and "will user use the system?"

Practitioners who desire to produce applications that have a higher probability of user acceptance can use the Usability-Acceptance Model to assist them in the development process. Specifically, when creating a user interface during the iterative design phase of an application development process, developers could measure user satisfaction and then modify the interface based on the results of the metric. Developers could infer from the variance explained of the Usability-Acceptance Model that improvements in the usability of the user interface, which is determined by increases in user satisfaction, would result in a higher probability of user acceptance once the applications is deployed.

The findings regarding the target technology used in this study, SRS, imply that users in a typical Air Force office environment would use SRS for text processing even though they would probably experience a slight decease in productivity. However, considering that the average user only "slightly agreed" on intending to use the target technology, suggests that users are not entirely convinced that SRS would provided much utility. If the average user indicated that he or she "agreed" or better yet, "strongly agreed" to use SRS, office managers could more confidently expect that an investment in this technology would be worth the expense of deploying it in their organization.

Limitations and Areas for Further Research

The most notable limitation in this study is the size of the subject pool used to gather data. A larger subject pool would have increased the statistical power of the results and may have more accurately represented the demographic diversity of the population of end users. Another limitation was the closed laboratory environment where the subjects accomplished the tasks of the experiment. The majority of office space in the workplace is composed of cubicles that are susceptible to a wide variety of noises and do not provide much privacy for its occupants. Future research utilizing a cubical type environment may yield results with stronger practical implications.

Speech recognition technology affords many other capabilities in addition to text processing, which was the capability utilized in this research. Future research could also include the other capabilities of SRS such as using speech functions for controlling applications such as web browsers. Additionally, the text processing tasks designed for this study focused on using SRS for transcribing a preexisting document into a word processor. Other tasks that would require a user to generate an original document using unstructured dictation might reveal different usability aspects of SRS. For example, task the user to compose a paragraph using SRS that describes a familiar, predetermined object in the room.

Conclusion

The results of this study failed to show that the P^3 model is an improved model for predicting user acceptance, as compared to TAM. Additionally, the objective usability metrics also failed to show a significant correlation to the question, "will users use the

system?" However, an alternate model for predicting user acceptance, the Usability-Acceptance Model (Figure 10), did emerge from the research which showed that the usability metric of user satisfaction can explain 53% of the variance of the MIS construct of a user's behavioral intention to use a new technology. The Usability-Acceptance Model is a first step towards bridging the important conceptual distinction between the questions of, "*can* users user the system?" and "*will* user use the system?" Last, although the results of this study indicate that users in a typical Air Force office environment would utilize SRS for text processing, the issue of increased productivity bears further study.

Appendix 1. Session 1 Experiment Script

PHASE I: INTRODUCTION

(Upon arrival, escort subject into SRS Lab, seat them at the computer and explain the experiment.)

Thank you for participating in this study. We are conducting research to evaluate how speech recognition software performs in comparison to conventional text input modes of keyboard and mouse. Before we get started, I'd like you to fill out this survey.

(After Subject completes survey)

As you know, you have been scheduled for two separate days of experimentation.

Today, day one of the experiment, you will train the speech recognition software to recognize your voice. Then, we will train you how to use the software.

On day two, we will review what we covered today, then you will perform timed tasks using both speech recognition software and keyboard and mouse. Keep in mind our purpose is to test the software, not your ability to use the software.

After the experiment on day two, you will also complete a survey asking about your perceptions of the speech recognition software.

Though your participation is greatly appreciated, you have the right to, at any time, terminate your participation in this experiment.

Do you have any medical conditions such as a head cold or sinus infection that might affect the way your voice normally sounds?

(Reschedule experiment time with subject if necessary)

Do you have any questions at this point?

(Answer any questions)

Now, go ahead and put on the headset and position microphone as outlined in the microphone positioning instructions on the screen.

The next few screens will calibrate the software volume to your voice.

Click "Next".

Read the instructions on the screen, then click "Start Test" and begin reading the text.

(After Beep)

Click "Next".

Click "Finish".

Click "Next".

Read the instructions on the screen, then click "Run Training Program".

Read the instructions on the screen, then click "Continue".

Read the instructions on the top of the active screen, then when you are ready to begin recording, click "Record" and begin dictating.

You will be dictating for about 30 minutes to train the software to learn your voice. If at anytime during training, you want to pause to ask questions, cough, drink water, or just take a break, click the pause button on the next screen.

Do you have any questions at this point?

(Answer any questions)

Select "Dogbert's Top Secret Management Handbook", then click "Train Now".

(After training is completed and window pops up)

Click "Finish".

(Ask if subject needs a break)

Now we will run an automated introductory training program in order to introduce you to the basic commands you will be using to operate the speech recognition

software. Once this initial training is complete you will be given the opportunity to practice what you've learned.

Do you have any questions at this point?

(Answer any questions)

Click "View Quick Tour".

Maximize the screen by clicking the middle button on the top-right of the active window.

For the following several screens, simply read the instructions on the left of the screen. This will give you the basic idea of the topic being presented. Then click the "Play" button to get a demonstration of that topic. When complete, click the "Next" button.

When done reading click "Play". (Screen 2)

When complete, click the "Next" button twice.

When done reading click "Play". (Screen 4)

When complete, click the "Next" button three times.

Click "Play". (Screen 7)

Click "Next".

Click "Play". (Screen 8)

Click "Next" button three times.

Click "Play". (Screen 11)

Click "Next".

Click "Play". (Screen 12)

Click "Next".

Click "Play". (Screen 13)

Close the active window by clicking the "X" at the top right of the screen.

Click "Next".

Click "Finish".

We will now reboot the computer. Feel free to ask any questions and take a break if you like.

    Reboot the Computer
    Start Word
    Start NaturalWord
    Open appropriate user file.

PHASE II: HANDS ON TRAINING - SCRIPTED

You will now be given an opportunity to practice the principles of speech recognition software use that you just learned.

(Hand subject Demo Task 1)

Here is a short text selection. You can either use this document stand and position the document stand wherever it is most comfortable for you, or you can hold the document in your hand as you dictate.

Simply dictate this text into the document. You can correct any errors as you go along using the commands you learned during the tutorial. Remember, as you dictate, words may not appear on the screen right away. Just keep reading naturally, while maintaining an awareness of the words on the screen so you can notice any errors as they occur. Feel free to ask any questions, but before you do, be sure to turn off the microphone. Once you're finished you will be given a chance to proof read and correct this document.

Before you begin, remember a few basic concepts:
1. Remember to take your time and speak clearly, enunciating each word. For this portion of the experiment, you will not be timed.

2. Don't forget to dictate punctuation (DEMO – For example comma, don't forget to say "period" say the end of a sentence period.).

3. There may be occasions where, as you are dictating, the software enters a word you did not intend. For example, you may say the word "speech", but the computer might enter the word "peach". This is called a mis-recognition error. In this case you can correct this error by selecting the mis-recognized word using the "select" command, then dictating the correct word again.

4. Also, remember that if the incorrect words are entered, you can use the command "Scratch That" to remove the incorrect words. Then re-dictate the phrase, making sure to enunciate each word. You can keep repeating this command until all errors are corrected.

5. Lastly, the microphone may pick up my voice or background sound and write unwanted text on the screen. In the event that this happens, simply say "Scratch That" or select the unwanted word(s) with the mouse and say "Delete That".

Once you start dictating, feel free to ask any questions. Be sure to turn off the microphone using the microphone icon on the tool bar.

Do you have any questions about specific commands at this point?

(Answer any questions)

Put back on the headset and turn on the microphone using the microphone icon on the tool bar, and begin dictating.

(When subject finishes)

Now you can check the document and correct any errors using voice commands. As you do this, we will coach you along. As we coach you, you do not need to turn off the microphone unless you need to ask a question. When correcting an error, there may be an occasion where the software just won't cooperate. When this happens, only try to correct the error three times. If the software fails to make the correction after three attempts, simply move on.

(Hand subject memory aid)

Here is a sheet containing some of the basic speech recognition commands you will be using for this experiment. The sheet is divided into four columns. The first two columns indicate the type and reference number of each command. The third column indicates the commands you actually say into microphone. Those commands are italicized. The commands in brackets you must say. The commands in parentheses are optional. You will perform each command until you are confident with them.

First we'll do some general formatting.

Using the mouse, position the cursor at the beginning of any sentence within the middle of the paragraph.

Insert a new line using command #1.

Position the cursor at the beginning of any sentence within the middle of the paragraph.

Insert a new paragraph using command #2. (After Subject performs the command) Notice the difference between command 1 and 2. Command 2 adds two carriage returns and command 1 only adds one carriage return.

Indent the first line of one of the paragraphs by using command #3. Position in the cursor at the beginning of any paragraph and say command #3.

Position the cursor between two words and practice command #4 a few times.

Now we will go through the text editing commands.

Select any single word in the document using command #5.
(Instruct the "two-word +" technique)

Now you can select a group of words using the "Select Through" command, #6.
Select the words "Talking, through Significantly". (Point the words out if necessary)

Now you can select a specific character, word, or line using command #7.
Position the cursor in the word "recognition". Now say "Select Character". Now say "Select Word". Now say "Select Line". (Quiz Subject if necessary)

Command #8 is Select Again. This command selects the next instance of the word currently selected. Select the word "recognition". Now select another instance of that word using command #8. Try this command two or three times. (Wait for Subject to complete the command two or three times) Notice how the software searches up from the bottom of the viewing area to select the next word.

Command #9 is Scratch That. We've practiced that already. Do you feel confident with that one?

Command #10, Delete That, is similar to scratch that. Simply select any word, or series of words, and say command #10, and it will be deleted. Try that now.

Command #10 is also good for the deleting extra spaces before or after words. Now create some extra spaces by using command #4 then use command #10 to delete the extra spaces you don't want.

Command #11 allows you do delete a number of characters or words in relation to the position to the cursor. Position the cursor somewhere within a paragraph, and delete the four previous words by saying "Delete last 4 words". Now delete the next 5 words by saying "Delete 5 Words". You can delete individual characters by saying "Characters" instead of "Words". Now try that.

Now, position the cursor somewhere between two words and try command #12. It's just like hitting the backspace key.

Commands #13, 14, and 15 allow you to copy, cut and paste text you selected. Go ahead and select any word on the document by using the select command. Say command #13 (copy). Position the cursor between any two words and say #15 (paste).

Now do the same procedure for command #14 (cut). Select any word on the document using the select command. Say command #14. Position the cursor between any two words and say command #15 (paste).

Command #16 is Undo That. That is just like hitting the undo button on the tool bar. It undoes the last command you dictated. Delete a word and then use command #16 to undo deletion.

Now we will move into text formatting commands.

Go ahead and select any word, or group of words, using the select command.

Now use commands #17, 18, and 19 to format them. (Coach Subject as necessary)

Once you've formatted a word, you can repeat the command to unformat the selected word. Try that.

These commands can also be used to format text before it is ever entered into the document. Position the cursor anywhere in the document using the mouse, say command #17, notice how of the bold icon on the toolbar turns on. (Point to bold icon if necessary)

Now say "speech recognition". (Pause) now say command #17 again to turn the bold function off. Notice how the bold icon on the toolbar turns off.

Command #20 capitalizes the first letter of the selected word or words. Select a word and capitalize it by saying "Cap That". Try that.

To remove capitalization, select the words and say "No Caps That". Try that.

These commands can also be used to capitalize a phrase you just said. Now dictate the phrase "capitalize a phrase you just said". Now use command #20 to capitalize the first letter of every word in the phrase. Now turn capitalization off by saying "No Caps That".

Go-ahead and turn the microphone off.

Do you have any questions about specific commands at this point?

(Answer any questions)

PHASE II: HANDS ON TRAINING - COACHED

In this phase of the training, you will enter text exactly as it appears on the hand out that I will give you using your voice and mouse. Feel free to use the full functionality of the mouse to select words, phrases and navigate throughout the document; however, you are not allowed to use the keyboard at all. **DO NOT** use the mouse to perform any text editing or formatting commands, use speech commands instead. As you enter text, feel free to correct any errors as they occur. I will coach you on the use of the speech recognition commands you just learned as you progress through the hand out. Feel free to stop dictating and ask questions at any time, but remember to turn off the microphone.

Here is a copy of the text selection [HAND SUBJECT DEMO TASK 2] you will enter into the computer using Microsoft Word. Feel free to use the memory aid to help you recall the speech recognition commands you learned earlier.

Do you have any questions at this point?

(Answer any questions)

Remember there will be a slight delay from the time you begin speaking to the time the software starts writing the text to the screen, just keep talking naturally and clearly and the computer will catch up to you. Remember, when you are trying to correct a mis- recognition, only give the software three chances and move on.

When you have completed entering the text, turn off the microphone by clicking the microphone icon on the tool bar and say "finished".

Now go ahead and turn on the microphone using the microphone icon on the tool bar.

When you are ready, begin speaking.

(Coach subject as required through the task)

(When subject is finished)

Do you have any questions?

(Answer any questions)

This completes the formal part of the training session. You may now practice any of the things you learned today or, if you feel comfortable with the software, you may leave.

Do you wish to practice a little now or you feel comfortable with the software?

(Thank the subject and remind subject to refrain from discussing details of the experiment to any potential subjects.)

(Remind subject of date and time of Day 2 - End Session 1)

Appendix 2. Session 2 Experiment Script - Manual Mode First

TASK B

Welcome back. You will recall on day one we trained the speech recognition software to recognize your voice. In addition, you learned how to use the software. Today you will perform two timed tasks using speech recognition software, keyboard and mouse. After each task you will be given an opportunity to proof your work. The proofing will also be timed. One task will involve direct text entry and formatting from a paper copy we will provide you using the speech recognition software commands combined with some mouse functionality. The other task will involve text entry and formatting using only the keyboard and mouse. I will give you detailed instructions for each task later.

Keep in mind our purpose is to test the software, not your ability to use the software. Although, your participation in this experiment greatly appreciated, you have the right to, at any time, terminate your participation in this experiment. Be assured that by participating in this experiment you will not be exposed to any risks. Also, we will keep the results of this experiment confidential and we will not release your name or your results to anyone.

At this point, do you have any questions?

(Answer any questions)

For Keyboard (B-1)

Open file "C:\MyDocuments\Subject Pool\Subject## \Subject ##-B1-Key.doc"

In this task you will enter the text from the selection exactly as it appears on the hand out that I will give you using your keyboard and mouse only. Any text formatting should be done as you go. You will not be using speech recognition at this point. Once you begin typing I will begin timing. As you type, feel free to correct any minor errors as they occur, realizing that you will be given an opportunity to more thoroughly proofread the document when you are finished. Before you begin typing, I will give you an opportunity to ask any questions. However, once you begin typing do not stop until you complete entering and formatting the text from the hand out.

Here is a copy of the text selection [HAND SUBJECT DOCUMENT B-1] you will enter into the computer using Microsoft Word.

Do you have any questions at this point?

(Answer any questions)

When you are done say "finished". Ready?

Begin.

(Start timing when subject starts typing)
(When subject is finished, stop timing and record time)

At this time, I need to prepare the computer for the next part of the first task. You can take a break if you like.

Save file
Save file as "C:\MyDocuments\SubjectPool\Subject##\Subject##-B1-Key-Proof.doc"

(When subject returns and is seated and ready)

You will now be given an opportunity to proof read and correct your work. Simply read the document to yourself, make sure the text and format were exactly entered as it appears on the hand out. Correct any errors you might find using your keyboard and mouse.

At this point, do you have any questions?

(Answer any questions)

When you are done say "finished". Ready?

Begin.

(Start timing when subject starts typing)
(When subject is finished, stop timing and record time)

Save and close file

Refresher Training

Before we begin with the speech recognition portion of the experiment, we will review some of the techniques and commands you learned on day one.

(Hand subject DEMO TASK 3)

Here is a short text selection. As on day one, simply enter the text as it appears on this sheet.

(Hand subject Memory Aid)

While completing this task, you can refer to this memory aid we went over on day one. Before you begin, are there any questions you have concerning specific commands?

(Answer any questions)

Before you begin, remember a few basic concepts:

1. **Remember to take your time and speak clearly, enunciating each word. For this portion of the experiment, you will not be timed.**

2. **Don't forget to dictate punctuation (DEMO – For example comma don't forget to say "period" say the end of a sentence period).**

3. **There may be occasions where, as you are dictating, the software enters a word you did not intend. For example, you may say the word "speech", but the computer might enter the word "peach". This is called a mis-recognition error. In this case you can correct this error by selecting the mis-recognized word using the "select" command, then dictating the correct word again.**

4. **Also, remember when you are correcting an error, only give the software three chances to fix the problem. If the software fails to make the correction after three attempts, move on. (Select [Char or Word or Line], Synonyms, Delete Cmnds)**

5. **Lastly, the microphone may pick up my voice or background sounds and write unwanted text on the screen. In the event that this happens, simply say "Scratch That" or select the unwanted word(s) with the mouse and say "Delete That".**

Once you start dictating, feel free to ask any questions. Remember to turn off the microphone before you do. Again, you will not be timed for this task. Do you have any question before you begin?

(Answer any questions and then signal Subject to start the task)

(Coach subject as required)

For Multi-Modal (B-2)

Open file "C:\MyDocuments\Subject Pool\Subject##\Subject##-B2-MM.doc"

Now we are ready to begin the speech recognition portion of the experiment. Now you will enter text exactly as it appears on the hand out that I will give you using your voice and mouse. Any text formatting should be done as you go. For this task, feel free to use the full functionality of the mouse to select words, phrases and navigate throughout the document. Do not use the mouse to perform any text editing or formatting commands. Once you begin dictating, I will begin timing. As you enter text, feel free to correct any minor errors as they occur, realizing that you will be given an opportunity to more thoroughly proofread the document when you are finished. Before you begin dictating, I will give you an opportunity to ask any questions. However, once you begin do not stop until you complete entering and formatting the text from the hand out. Also, I will not be able to assist you during this task.

Here is a copy of the text selection [HAND SUBJECT DOCUMENT B-2] you will enter into the computer using Microsoft Word. Also, remember to use the memory aid to help you recall the speech recognition commands you learned on day one.

Do you have any questions at this point?

(Answer any questions)

Remember there may be a slight delay from the time you begin speaking to the time the software starts writing the text to the screen, just keep talking naturally and clearly and the computer will catch up to you.

Also, remember when you are correcting an error, only give the software three chances to fix the problem. If the software fails to make the correction after three attempts, move on.

When you have completed entering the text, turn off the microphone by clicking the microphone icon on the tool bar and say "finished".

Now go ahead and turn on the microphone using the microphone icon on the tool bar.

When you are ready, begin speaking.

(Start timing when subject starts speaking)
(When subject is finished, stop timing and record time)

At this time, we need to prepare the computer for the next part of the task. You can take a quick break if you like.

Save file
Save file as "C:\MyDocuments\SubjectPool\Subject##\Subject##-B2-MM-Proof.doc"

(When subject returns and is seated and ready)

You will now be given an opportunity to proof read and correct your work. Simply read the text on the screen to yourself, make sure the text and format was entered exactly as it appears on the hand out. You may use the speech recognition commands on the memory aid to correct the document. Feel free to use the full functionality of the mouse to select words, phrases and navigate throughout the document. Do not use the mouse to perform any text editing or formatting commands.

At this point, do you have any questions?

When you're finished proofing the text, turn off your microphone by clicking the microphone icon on the tool bar and then say "finished".

At this point, do you have any questions?

(Answer any questions)

Now go ahead and turn on the microphone using the microphone icon on the tool bar.

When you are ready, begin speaking.

(Start timing when subject starts speaking)
(When subject is finished, stop timing and record time)

Save close file

(Administer Exit Survey)

(After Subject turns in survey)

<u>Outbrief</u>
- **Review times and observations with subject**
- **Don't discuss**
- **We will keep results confidential**
- **Thanks again!**

Appendix 3. Session 2 Experiment Script - SRS Mode First

TASK B

Welcome back. You will recall on day one we trained the speech recognition software to recognize your voice. In addition, you learned how to use the software. Today you will perform two timed tasks using speech recognition software, keyboard and mouse. After each task you will be given an opportunity to proof your work. The proofing will also be timed. One task will involve direct text entry and formatting from a paper copy we will provide you using the speech recognition software commands combined with some mouse functionality. The other task will involve text entry and formatting using only the keyboard and mouse. I will give you detailed instructions for each task later.

Keep in mind our purpose is to test the software, not your ability to use the software. Although, your participation in this experiment greatly appreciated, you have the right to, at any time, terminate your participation in this experiment. Be assured that by participating in this experiment you will not be exposed to any risks. Also, we will keep the results of this experiment confidential and we will not release your name or your results to anyone.

At this point, do you have any questions?

(Answer any questions)

<u>Refresher Training</u>
Before we begin with the speech recognition portion of the experiment, we will review some of the techniques and commands you learned on day one.

(Hand subject DEMO TASK 3)

Here is a short text selection. As on day one, simply enter the text as it appears on this sheet.

(Hand subject Memory Aid)

While completing this task, you can refer to this memory aid we went over on day one. Before you begin, are there any questions you have concerning specific commands?

(Answer any questions)

Before you begin, remember a few basic concepts:
1. Remember to take your time and speak clearly, enunciating each word. For this portion of the experiment, you will not be timed.

2. Don't forget to dictate punctuation (DEMO – For example comma don't forget to say "period" say the end of a sentence period).

3. There may be occasions where, as you are dictating, the software enters a word you did not intend. For example, you may say the word "speech", but the computer might enter the word "peach". This is called a mis-recognition error. In this case you can correct this error by selecting the mis-recognized word using the "select" command, then dictating the correct word again.

4. Also, remember when you are correcting an error, only give the software three chances to fix the problem. If the software fails to make the correction after three attempts, move on. (Select [Char or Word or Line], Synonyms, Delete Cmnds)

5. Lastly, the microphone may pick up my voice or background sounds and write unwanted text on the screen. In the event that this happens, simply say "Scratch That" or select the unwanted word(s) with the mouse and say "Delete That".

Once you start dictating, feel free to ask any questions. Remember to turn off the microphone before you do. Again, you will not be timed for this task. Do you have any question before you begin?

(Answer any questions and then signal Subject to start the task)

(Coach subject as required)

For Multi-Modal (B-1)

Open file "C:\MyDocuments\Subject Pool\Subject##\Subject##-B1-MM.doc"

Now we are ready to begin the speech recognition portion of the experiment. Now you will enter text exactly as it appears on the hand out that I will give you using your voice and mouse. Any text formatting should be done as you go. For this task, feel free to use the full functionality of the mouse to select words, phrases and navigate throughout the document. Do not use the mouse to perform any text editing or formatting commands. Once you begin dictating, I will begin timing. As you enter text, feel free to correct any minor errors as they occur, realizing that you will be given an opportunity to more thoroughly proofread the document when you are finished. Before you begin dictating, I will give you an opportunity to ask any questions. However, once you begin do not stop until you complete entering and formatting the text from the hand out. Also, I will not be able to assist you during this task.

Here is a copy of the text selection [HAND SUBJECT DOCUMENT B-1] you will enter into the computer using Microsoft Word. Also, remember to use the memory aid to help you recall the speech recognition commands you learned on day one.

Do you have any questions at this point?

(Answer any questions)

Remember there may be a slight delay from the time you begin speaking to the time the software starts writing the text to the screen, just keep talking naturally and clearly and the computer will catch up to you.

Also, remember when you are correcting an error, only give the software three chances to fix the problem. If the software fails to make the correction after three attempts, move on.

When you have completed entering the text, turn off the microphone by clicking the microphone icon on the tool bar and say "finished".

Now go ahead and turn on the microphone using the microphone icon on the tool bar.

When you are ready, begin speaking.

(Start timing when subject starts speaking)
(When subject is finished, stop timing and record time)

At this time, we need to prepare the computer for the next part of the task. You can take a quick break if you like.

Save file
Save file as "C:\MyDocuments\SubjectPool\Subject##\Subject##-B1-MM-Proof.doc"

(When subject returns and is seated and ready)

You will now be given an opportunity to proof read and correct your work. Simply read the text on the screen to yourself, make sure the text and format was entered exactly as it appears on the hand out. You may use the speech recognition commands on the memory aid to correct the document. Feel free to use the full functionality of the mouse to select words, phrases and navigate throughout the document. Do not use the mouse to perform any text editing or formatting commands.

At this point, do you have any questions?

When you're finished proofing the text, turn off your microphone by clicking the microphone icon on the tool bar and then say "finished".

At this point, do you have any questions?

(Answer any questions)

Now go ahead and turn on the microphone using the microphone icon on the tool bar.

When you are ready, begin speaking.

(Start timing when subject starts speaking)
(When subject is finished, stop timing and record time)

Save close file

For Keyboard (B-2)

Open file "C:\MyDocuments\Subject Pool\Subject##\Subject##-B2-Key.doc"

In this task you will enter the text from the selection exactly as it appears on the hand out that I will give you using your keyboard and mouse only. You will not be using speech recognition at this point. Once you begin typing I will begin timing. As you type, feel free to correct any minor errors as they occur, realizing that you will be given an opportunity to more thoroughly proofread the document when you are finished. Before you begin typing, I will give you an opportunity to ask any questions. However, once you begin typing do not stop until you complete entering and formatting the text from the hand out.

Here is a copy of the text selection [HAND SUBJECT DOCUMENT B-2] you will enter into the computer using Microsoft Word.

Do you have any questions at this point?

(Answer any questions)

When you are done say "finished". Ready?

Begin.

(Start timing when subject starts typing)
(When subject is finished, stop timing and record time)

At this time, I need to prepare the computer for the next part of the first task. You can take a break if you like.

Save file
Save file as "C:\MyDocuments\SubjectPool\Subject##\Subject##-B2-Key-Proof.doc"

(When subject returns and is seated and ready)

You will now be given an opportunity to proof read and correct your work. Simply read the document to yourself, make sure the text and format were exactly entered as it appears on the hand out. Correct any errors you might find using your keyboard and mouse.

At this point, do you have any questions?

(Answer any questions)

When you are done say "finished". Ready?

Begin.

(Start timing when subject starts typing)
(When subject is finished, stop timing and record time)

Save and close file

(Administer Exit Survey)

(After Subject turns in survey)

<u>Outbrief</u>

       - Review times and observations with subject

       - Don't discuss

       - We will keep results confidential

       - Thanks again!

Appendix 4. SRS Initial Training Task

Recent advances in the area of low-cost speech recognition have moved the technology into everyday consumer products. With vast improvements in the quality and accuracy of cheap speech recognition systems, the value of adding speech recognition technology to everyday customer products is now being realized. As products become increasingly complicated and offer more functions, implementing speech recognition allows consumers to use products more intuitively while maximizing their functionality. Talking to our products and listening to what they say gives products a life of their own and significantly changes the way we can use them. One area which speech recognition will have a deep impact is voice dialing. This allows a consumer to dial a phone number simply by saying the name of the person they wish to call. Sensory Inc. first made its reputation as a company that made toys talk.

DEMO TASK 1

Appendix 5. SRS Memory Aid

TYPE	#	SPEECH COMMAND	EXPLINATION OF COMMANDS
General Formatting	1	*New Line*	Adds one carriage return and does not automatically capitalized the next word you dictate.
	2	*New Paragraph*	Adds two carriage returns and capitalizes the next word you dictate.
	3	*Tab Key*	Moves the insertion point to the next tab stop.
	4	*Space Bar*	Adds one space where the cursor is positioned.
Text Editing	5	*Select <text>*	<text> means currently visible, contiguous words.
	6	*Select <start-text> Through <end-text>*	<start-text> means the first word you want to select. <end -text> means the last word you want to select.
	7	*Select [Character or Word or Line]*	Selects character, word, line or paragraph based on the position of the cursor.
	8	*Select Again*	Selects the next instance of the word currently selected.
	9	*Scratch That*	Deletes selected text or the last word you say. You can repeat Scratch That up to 10 times.
	10	*Delete That*	Deletes selected text, the next character after the cursor or the last word you said.
	11	*Delete [Next or Last] # [Character(s) or Word(s)]*	Deletes a specified number of characters or words forward from the cursor or back from the cursor.
	12	*Press Backspace*	Same functionality as pressing the "Backspace" the key.
	13	*Copy That*	Copies selected text.
	14	*Cut That*	Cuts selected text.
	15	*Paste That*	Pastes selected text.
	16	*Undo That*	Undoes last action.
Text Formatting	17	*Bold That*	Bolds selected text or the word you said last.
	18	*Underline That*	Underlines selected text or the word you said last.
	19	*Italicize That*	Italicizes selected text or the words you said last.
	20	*Cap That or No Caps That*	Capitalizes first letter of the text you selected or removes all capitalization of selected text.

Appendix 6. SRS Supplemental Training Task

1.1.3.1 Microprocessors

The dramatic and continuing growth in the speed and power of microprocessors is a primary factor in the migration of **advanced speech** recognition technology from laboratories to real-world applications. Figure 1.1 provides a **dramatic example** of that growth for microprocessors. The advent of each new generation of chips has heralded the commercialization of a new, more advanced class of speech recognition <u>systems</u> and <u>technology</u>.

1.1.3.2 The Effects of Miniaturization

One measure of **progress** is the increasing number of components we can cram onto a silicon chip about the size of a fingernail. For signal-processing chips, the scale of integration is about 33 percent per year. At the same time, the speed of individual components is increasing about 20 percent each year.

Miniaturization of hardware is fostering the use of <u>**speech recognition**</u> in consumer products. As smaller systems become more powerful, they can support increasingly complex speech-recognition technology. Several small-vocabulary, chip based systems were introduced in the early 1990s.

DEMO TASK 2

Appendix 7. SRS Experimental Task B-1

1.1.2 The Challenge

The success of *speech recognition* software resulted from 40 years of work that has produced technology capable of accurately processing spoken input containing sizable vocabularies. Despite those impressive achievements, *speech recognition* still has not reached its goal of developing systems capable of understanding virtually anything anyone says on any topic when they are speaking in a natural free-flowing style of speech and situated in almost any speaking environment, no matter how noisy, that is, to understand spoken language as well as humans can. This shortcoming may be surprising since, for many humans, understanding what other people say may seem to be a simple task. In fact, it is extremely complex and difficult. There are many reasons why.

1. Voluminous Data

Although it may seem as if we speak using a single tone, the quantity of data in the sound wave is overwhelming. Within the range of human hearing, speech sounds can span more than **20,000 frequencies**.

- The time required to capture, digitize, and recognize frequency patterns for every fraction of a second of speech would overwhelm any PC on the market as well as most other computer systems.

- In order to recognize speech at a speed that is acceptable to users, the amount of data and the signal must be dramatically reduced. It is not necessary to manipulate all the data from the entire speech wave. Some excludable data are irrelevant to the recognition process while other pieces of data are redundant.

- The quantity of the data can be reduced further by taking samples from the signal rather than trying to process the entire waveform.

2. Sound Wave

The paucity of information in the speech sound wave may appear to contradict the preceding point, but it simply highlights the fact that speech is more than acoustic sound patterns. Spoken language interaction between people requires knowledge about **word meanings, communication patterns, and the world in general**. Words with widely different meanings and usage patterns may share the same sequence of sound patterns. These are sets of frequently occurring words that sound the same but are spelled differently. Using these systems, the meanings of words that affect the interpretation of utterances cannot be extracted from the sound stream alone.

Often, a grammar is required to assist in the process. A grammar that links appropriate words using distinctions would, for example, link words in a way that makes sense. Knowledge of the world would be needed to determine the correct meaning of each sentence. Similar examples requiring world knowledge that is currently unavailable to computers can be drawn from newspaper headlines. Fortunately, the inclusion of information beyond acoustic analysis of the sound stream is not needed for many simple applications. It is obvious, though, that the incorporation of such "**higher level**" knowledge into a speech recognition system would serve as a gateway to truly natural speech communication with machines.

3. Speech Flow

Since we speak in individual words and we "hear" what other people say as sequences of words, it seems reasonable to expect the speech sound wave to consist of words with clearly marked boundaries. Unfortunately, that is not at all the case. Speech is uttered as a continuous flow of sounds and even when words are spoken distinctly there are no inherent separations between them. This should not be surprising since we hear foreign languages as streams of sound unbroken by our recognition of distinct words. The same phenomenon occurs for unfamiliar **words and phrases** in English. Once it moved beyond single word input, speech recognition was forced to address the problem of segmenting the speech stream into its component words.

Appendix 8. SRS Experimental Task B-2

4. Variability

One person's voice and **speech patterns** can be entirely different from those of another person. Some elements of this diversity are physical. Each individual is unique, differing from others in the size and shape of their mouths, the length and width of their necks, and a range of other physical characteristics. Added to these anatomical variations are age, sex, regional dialect, health, and an individual's **personal style** of speech. Despite these differences, a recognition system must be able to accurately process the speech of anyone who is expected to use the speech system. The development of speaker modeling techniques has produced dramatic advances in handling inter-speaker variability. Technologies alone will not eliminate all of these issues. Resolution of a significant portion of speaker variability issues, including speaker training, vocabulary selection, and the human factors in application design, all affect the ability of a recognition system to handle inter-speaker variability. These concerns are the responsibility of application designers.

5. More Variability

Even a single speaker will exhibit variability. The sound pattern of a word changes when speakers whisper or shout, when they are angry or sad, and when they are tired or ill.

- Even when speaking **normally**, individual speakers rarely say a word the same way twice.

- In fact, variability is the basic characteristic of speech. When speaker variability is added to inter-speaker differences it becomes difficult to identify and extract critical, word identification information from the input.

- Speaker modeling techniques have been designed to extract common intra-speaker patterns of the variability and produce very high **speech recognition** accuracy.

6. Noise

Natural speaking environments bombard the speaker with sounds of varying wildness emanating from many sources. They include people speaking in the background, street sounds, the slam of a door, music, and the loud noise of machinery. Sometimes the noise in a speaking environment can be so great that people cannot understand each other. As speech recognition is embedded in more diverse products and systems, the spectrum of noises will also grow. **Unfortunately, the challenging speaking environments are the ones that most characterize our daily living: busy offices, factories, loading docks, airports, automobiles,** and even our own homes. Background noise is not the only intrusion speech recognition systems must combat.

They must handle noise produced by the input device, sounds made by the speaker, such as lip smacks, and non-communication vocal limitations made by the speaker. Speech recognition over telephones is becoming increasingly popular, but it is one of the most challenging of speaking channels. Even people have trouble with it. Voices can be faint or full of static, but when everything is functioning well, it may still be difficult to distinguish between similar sounding words and sounds. The success of speech recognition over the telephone illustrates the recent progress that has been made in this area. As with other issues, the role of the application developer in addressing noise has a strong impact on the ultimate success of the speech recognition application. The **rapid technological** advances of the last 15 years have come far toward achieving those goals, but the challenge should not be underestimated.

1.1.3 Driving Forces

Speech recognition has only recently achieved a level of reliability and flexibility to attract the interest of business and consumers.
Its achievements are due, in part, to significant technological advances within the industry. **Equally important** are external factors that have functioned as driving forces for speech recognition.

Appendix 9. SRS Practice Task

1.1.3.3 Global Business

The world is growing smaller politically and economically. **International business ventures** that link professionals on opposite sides of the globe are becoming commonplace. This has spawned a need to establish 24-hour telecommunications capabilities. Some of these needs can be satisfied by hiring "*bilingual*" telephone operators and business professionals. That solution is not always necessary or affordable, and touch-tone technology is not widely available outside of **North America**.

1.2 Historical Overview

The first documented attempts to construct an automatic speech recognition system occurred long before the **digital computer** was invented. In the 1870s Alexander Graham Bell wanted to build a device that would make speech visible to hearing-impaired people. He ended up inventing the telephone. Many years later, a Hungarian scientist requested permission for a patent to develop an automatic transport system using the optical sound tracks of movie films. The soundtrack was to serve as a source of capturing the sound patterns of speech. The system would identify the sound sequences and print them out. The request for a patent was labeled unrealistic and denied.

DEMO TASK 3

Appendix 10. SRS Usage Perception Questionnaire

Questions 1 - 26 will ask you about your general perceptions of speech recognition software. Please read each of the following statements and check appropriate box that best reflects your opinion or completes the statement.

1. Learning to operate Dragon Speech Recognition Software (SRS) was easy for me.
□ Strongly Agree (1)   □ Agree (2)   □ Slightly Agree (3)   □ Neutral (4)   □ Slightly Disagree (5)   □ Disagree (6)   □ Strongly Disagree (7)

2. Using Dragon SRS would improve my performance at work.
□ Strongly Agree (1)   □ Agree (2)   □ Slightly Agree (3)   □ Neutral (4)   □ Slightly Disagree (5)   □ Disagree (6)   □ Strongly Disagree (7)

3. I found it easy to get Dragon SRS to do what I wanted it to do.
□ Strongly Agree (1)   □ Agree (2)   □ Slightly Agree (3)   □ Neutral (4)   □ Slightly Disagree (5)   □ Disagree (6)   □ Strongly Disagree (7)

4. I found Dragon SRS enjoyable to use.
□ Strongly Agree (1)   □ Agree (2)   □ Slightly Agree (3)   □ Neutral (4)   □ Slightly Disagree (5)   □ Disagree (6)   □ Strongly Disagree (7)

5. Using Dragon SRS is a(n)_____idea.
□ Extremely Good (1)   □ Good (2)   □ Slightly Good (3)   □ Neither Good nor Bad (4)   □ Slightly Bad (5)   □ Bad (6)   □ Extremely Bad (7)

6. It would be easy for me to become skillful using Dragon SRS.
□ Strongly Agree (1)   □ Agree (2)   □ Slightly Agree (3)   □ Neutral (4)   □ Slightly Disagree (5)   □ Disagree (6)   □ Strongly Disagree (7)

7. Using Dragon SRS would enhance my effectiveness at work.
□ Strongly Agree (1)   □ Agree (2)   □ Slightly Agree (3)   □ Neutral (4)   □ Slightly Disagree (5)   □ Disagree (6)   □ Strongly Disagree (7)

8. Using Dragon SRS is a(n)_____idea.
□ Extremely   □ Foolish   □ Slightly   □ Neither Wise   □ Slightly   □ Wise   □ Extremely

9. I would find Dragon SRS easy to use.

☐ Strongly   ☐ Agree   ☐ Slightly   ☐ Neutral   ☐ Slightly   ☐ Disagree   ☐ Strongly
   Agree (1)   (2)   Agree (3)   (4)   Disagree (5)   (6)   Disagree (7)

10. Using Dragon SRS would increase my productivity at work.

☐ Strongly   ☐ Agree   ☐ Slightly   ☐ Neutral   ☐ Slightly   ☐ Disagree   ☐ Strongly
   Agree (1)   (2)   Agree (3)   (4)   Disagree (5)   (6)   Disagree (7)

11. Dragon SRS is _____ to operate.

☐ Extremely   ☐ Fun   ☐ Slightly   ☐ Neither Fun   ☐ Slightly   ☐ Boring   ☐ Extremely
   Fun (1)   (2)   Fun (3)   nor Boring (4)   Boring (5)   (6)   Boring (7)

12. I_____the idea of using Dragon SRS.

☐ Strongly   ☐ Like   ☐ Slightly   ☐ Don't Care   ☐ Slightly   ☐ Dislike   ☐ Strongly
   Like (1)   (2)   Like (3)   About (4)   Dislike (5)   (6)   Dislike (7)

13. I would find Dragon SRS useful at work.

☐ Strongly   ☐ Agree   ☐ Slightly   ☐ Neutral   ☐ Slightly   ☐ Disagree   ☐ Strongly
   Agree (1)   (2)   Agree (3)   (4)   Disagree (5)   (6)   Disagree (7)

14. Using Dragon SRS is entirely within my control.

☐ Strongly   ☐ Agree   ☐ Slightly   ☐ Neutral   ☐ Slightly   ☐ Disagree   ☐ Strongly
   Agree (1)   (2)   Agree (3)   (4)   Disagree (5)   (6)   Disagree (7)

15. Using Dragon SRS was a pleasurable experience for me.

☐ Strongly   ☐ Agree   ☐ Slightly   ☐ Neutral   ☐ Slightly   ☐ Disagree   ☐ Strongly
   Agree (1)   (2)   Agree (3)   (4)   Disagree (5)   (6)   Disagree (7)

16. Using Dragon SRS would be_____.

☐   ☐ Pleasant   ☐ Slightly   ☐ Neither Pleasant   ☐ Slightly   ☐ Unpleasant   ☐ Extremely
Extremely   (2)   Pleasant   nor Unpleasant (4)   Unpleasant (5)   (6)   Unpleasant(7)
Pleasant (1)   (3)

17. People who influence my behavior would think that I should use Dragon SRS.

☐ Strongly   ☐ Agree   ☐ Slightly   ☐ Neutral   ☐ Slightly   ☐ Disagree   ☐ Strongly
   Agree (1)   (2)   Agree (3)   (4)   Disagree (5)   (6)   Disagree (7)

18. I would be able to use Dragon SRS.

☐ Strongly   ☐ Agree   ☐ Slightly   ☐ Neutral   ☐ Slightly   ☐ Disagree   ☐ Strongly

19. Dragon SRS was a tedious application to use.
☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

20. I have the knowledge and the ability to make use of Dragon SRS.
☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

21. People who are important to me would think that I should use Dragon SRS.
☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

22. I would use Dragon SRS at work if it was made available to me.
☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

23. I have the knowledge and ability to effectively use Dragon SRS at work.
☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

24. If I had a system capable of running Dragon SRS at work, I would use the software.
☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

25. I intend to recommend Dragon SRS to my co-workers based on my positive experience.
☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

26. I think using Dragon SRS is _____ .
☐ Extremely Amusing (1) ☐ Amusing (2) ☐ Slightly Amusing (3) ☐ Neither Dull nor Amusing (4) ☐ Slightly Dull (5) ☐ Dull (6) ☐ Extremely Dull (7)

NEXT SECTION

Questions 27 - 50 will ask you about your perceptions of the functionality of speech recognition software versus keyboard and mouse. Please read each of the following statements and check appropriate box that best reflects your opinion or completes the statement.

27. I _____ using Dragon SRS to input text rather than using a keyboard.

☐ Strongly Prefer (1) ☐ Prefer (2) ☐ Slightly Prefer (3) ☐ Neither Prefer nor Oppose (4) ☐ Slightly Oppose (5) ☐ Oppose (6) ☐ Strongly Oppose (7)

28. I prefer a mouse to format text (bold, <u>underline</u>, *italic*, etc.) instead of Dragon SRS.

☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

29. I would rather correct Dragon SRS misrecognition errors than my own mis-spellings.

☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

30. I prefer to select text using a mouse over speech commands.

☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

31. I would choose speech commands to format text before using a mouse to format text.

☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

32. It is _____ correcting Dragon SRS misrecognition errors than my mis-spelled words.

☐ Extremely Faster (1) ☐ Faster (2) ☐ Slightly Faster (3) ☐ Neither Faster nor Slower (4) ☐ Slightly Slower (5) ☐ Slower (6) ☐ Extremely Slower (7)

33. If given the option, I would choose Dragon SRS to enter text over a keyboard.

☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

34. It is *faster* using a mouse to select text than it is with speech commands.

☐ Strongly ☐ Agree ☐ Slightly ☐ Neutral ☐ Slightly ☐ Disagree ☐ Strongly

35. Entering text with a keyboard is more effective than entering text with Dragon SRS.

☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

36. It is _____ to format text with speech commands rather than with a mouse.

☐ Extremely More Effective (1) ☐ More Effective (2) ☐ Slightly Effective (3) ☐ Neither More nor Less Effective (4) ☐ Slightly Less Effective (5) ☐ Less Effective (6) ☐ Extremely Less Effective (7)

37. Using speech commands to select text is _____ than using a mouse.

☐ Extremely Efficient (1) ☐ More Efficient (2) ☐ Slightly Efficient (3) ☐ Neither More nor Less Efficient (4) ☐ Slightly Less Efficient (5) ☐ Less Efficient (6) ☐ Extremely Less Efficient (7)

38. I _____ correcting my mis-spellings instead of Dragon SRS misrecognition errors.

☐ Strongly Prefer (1) ☐ Prefer (2) ☐ Slightly Prefer (3) ☐ Neither Prefer nor Oppose (4) ☐ Slightly Oppose (5) ☐ Oppose (6) ☐ Strongly Oppose (7)

39. I think inputting text with Dragon SRS is more *effective* than with a keyboard.

☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

40. Using Dragon SRS to select text is more *effective* than using a mouse for text selection.

☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

41. Correcting mis-spelling errors is more *effective* than correcting misrecognition errors.

☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

42. Using a mouse to format text is more *efficient* than using speech commands.

☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

43. I do not see any gains in *efficiency* for myself using Dragon SRS for entering text.

☐ Strongly ☐ Agree ☐ Slightly ☐ Neutral ☐ Slightly ☐ Disagree ☐ Strongly

45. Formatting text with a mouse is more effective than formatting text with Dragon SRS.

☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

46. I believe using Dragon SRS for inputting text is more *efficient* to using a keyboard.

☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

47. Using Dragon SRS to format text is a(n) _____ method than using a mouse.

☐ Extremely More Efficient (1) ☐ More Efficient (2) ☐ Slightly Efficient (3) ☐ Neither More nor Less Efficient (4) ☐ Slightly Less Efficient (5) ☐ Less Efficient (6) ☐ Extremely Less Efficient (7)

48. Using a mouse to correct mis-spellings is more *efficient* than using Dragon SRS to
    correct misrecognition errors.

☐ Strongly Agree (1) ☐ Agree (2) ☐ Slightly Agree (3) ☐ Neutral (4) ☐ Slightly Disagree (5) ☐ Disagree (6) ☐ Strongly Disagree (7)

49. I think it is _____ using a mouse to select text than it is using speech commands.

☐ Extremely More Effective (1) ☐ More Effective (2) ☐ Slightly Effective (3) ☐ Neither More nor Less Effective (4) ☐ Slightly Less Effective (5) ☐ Less Effective (6) ☐ Extremely Less Effective (7)

50. I think it is _____ correcting misrecognition errors rather than my spellings errors.

☐ Extremely More Effective (1) ☐ More Effective (2) ☐ Slightly Effective (3) ☐ Neither More nor Less Effective (4) ☐ Slightly Less Effective (5) ☐ Less Effective (6) ☐ Extremely Less Effective (7)

NEXT SECTION

Questions 51 - 62 will ask you about your perceptions of the usability of speech recognition software. Please read each of the following statements and check appropriate box that best reflects your opinion or completes the statement.

Overall reaction to the software:

		0	1	2	3	4	5	6	7	8	9	
51.	Terrible	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Wonderful
52.	Difficult	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Easy
53.	Frustrating	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Satisfying
54.	Dull	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Stimulating
55.	Rigid	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Flexible

Exploring SRS features by trial and error:

		0	1	2	3	4	5	6	7	8	9	
56.	Difficult	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Easy

Remembering the names and use of speech commands:

		0	1	2	3	4	5	6	7	8	9	
57.	Difficult	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Easy

Performing speech commands are straightforward:

		0	1	2	3	4	5	6	7	8	9	
58.	Never	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Always

59. I feel comfortable using Dragon SRS.

☐ Strongly Agree (1)　　☐ Agree (2)　　☐ Slightly Agree (3)　　☐ Neutral (4)　　☐ Slightly Disagree (5)　　☐ Disagree (6)　　☐ Strongly Disagree (7)

60. I believe I became productive quickly using Dragon SRS.

□ Strongly   □ Agree   □ Slightly   □ Neutral   □ Slightly   □ Disagree   □ Strongly
  Agree (1)   (2)   Agree (3)   (4)   Disagree (5)   (6)   Disagree (7)

61. Whenever I make a mistake using Dragon SRS, I recover easily and quickly.

□ Strongly   □ Agree   □ Slightly   □ Neutral   □ Slightly   □ Disagree   □ Strongly
  Agree (1)   (2)   Agree (3)   (4)   Disagree (5)   (6)   Disagree (7)

62. Dragon SRS has all the functions and capability I expect it to have.

□ Strongly   □ Agree   □ Slightly   □ Neutral   □ Slightly   □ Disagree   □ Strongly
  Agree (1)   (2)   Agree (3)   (4)   Disagree (5)   (6)   Disagree (7)

63. Overall, I am satisfied with Dragon SRS.

□ Strongly   □ Agree   □ Slightly   □ Neutral   □ Slightly   □ Disagree   □ Strongly
  Agree (1)   (2)   Agree (3)   (4)   Disagree (5)   (6)   Disagree (7)

NEXT SECTION

Factor to Question Index

	EOU	U	A	BI
Questions	1, 3, 6, 9	2, 7, 10, 13	5, 8, 12, 16	22, 24

	Intrinsic Motivation	
Questions	4, 11, 15, 19, 25, 26	

Usability			
Categories	Text Entry	Text Formatting	Text Correction	Text Selection
Preference	27, 33	28, 31	29, 38	30, 44
Perceived Effectiveness	35, 39	36, 45	41, 50	40, 49
Perceived Efficiency	43, 46	42, 47	32, 48	34, 37
	Overall Reaction (QUIS)	Learning (QUIS)	IBM Computer Usability Sat. Ques.	
Satisfaction	51-55	56-58	59-63	

Use this last page to write any comments you may have about your perceptions of speech recognition software, its functionality compared to keyboard and mouse, or anything else you would like to comment on that was or was not addressed in the survey. Thanks again!

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

(Use Back Of Sheet If Necessary)

Factor to Question Index

	EOU	U	A	BI
Questions	1, 3, 6, 9	2, 7, 10, 13	5, 8, 12, 16	22, 24

	Intrinsic Motivation	
Questions	4, 11, 15, 19, 25, 26	

Usability				
Categories	Text Entry	Text Formatting	Text Correction	Text Selection
Preference	27, 33	28, 31	29, 38	30, 44
Perceived Effectiveness	35, 39	36, 45	41, 50	40, 49
Perceived Efficiency	43, 46	42, 47	32, 48	34, 37
	Overall Reaction (QUIS)	Learning (QUIS)	IBM Computer Usability Sat. Ques.	
Satisfaction	51-55	56-58	59-63	

Appendix 11. Pre-experiment Questionnaire

Subject#: _____

1. Name:_____

Daytime Phone:_____ E-Mial:_____

2. Sex (Circle 1): M or F

3. Age:_____

4. Rank (Check 1):
   - ☐ Airman (E1 to E4)   ☐ CGO (O1 to O3)
   - ☐ NCO (E5 and E6)   ☐ FGO (O4 to O6)
   - ☐ Sr. NCO (E7 to E9)   ☐ Civilian

5. Proficiency with MS Word (Check 1)
   - ☐ None
   - ☐ Novice
   - ☐ Intermediate
   - ☐ Advanced

6. Approximately, how many hours do you use MS Word per week? (Check 1):
   - ☐ None
   - ☐ 1 - 5 hours
   - ☐ 5 - 10 hours
   - ☐ 10 - 20 hours
   - ☐ More than 20 hours

7. Proficiency with speech recognition software (Check 1):
   - ☐ None
   - ☐ Novice
   - ☐ Intermediate
   - ☐ Advanced

8. Proficiency with Dragon NaturallySpeaking speech recognition software (Check 1):
   - ☐ None
   - ☐ Novice
   - ☐ Intermediate
   - ☐ Advanced

Bibliography

Adams, Dennis A., Ryan R. Nelson and Peter A. Todd. "Perceived Usefulness, Ease of Use, and usage of Technology: A Replication," <u>MIS Quarterly, 16-2:</u> 225-247 (1992).

Agarwal, Ritu and Jayesh Prasad. "The Role of Innovation Characteristics and Perceived Voluntariness in the Acceptance of Information Technologies," <u>Decision Sciences, 28-3:</u> 557-582 (1997).

Ajzen, Izak. "The Theory of Planned Behavior," <u>Organizational Behavior and Human Decision Processes, 50:</u> 179-211 (1991).

Ajzen, I. and M. Fishbein. <u>Understanding Attitudes and Predicting Social Behavior.</u> Prentice-Hall, Englewood Cliffs, NJ: 1980.

Alwang, Greg. "Speech Recognition: Finding Its Voice," <u>PC Magazine, 17-18:</u> 191-204 (20 October 20 1998).

Anderson, Nancy, S. and Judith R. Olsen. "Methods for Designing Software to Fit Human Needs and Capabilities," in <u>Readings in Human-Computer Interaction: A Multidisciplinary Approach.</u> Ronald M. Baecker and William A. S. Buxton. Los Altos, CA: Morgan Kaufmann, 1987.

Baecker, Ronald M., Jonathan Grudin, William A. S. Buxton, and Saul Greenberg. <u>Readings in Human-Computer Interaction: Toward the Year 2000.</u> San Francisco, CA: Morgan Kaufmann, 1995.

Bannon, Liam J. "Issues in Design: Some Notes," in <u>User Centered System Design: New Perspectives in Human-Computer Interaction.</u> Donald A. Norman and Stephen W. Draper. Hillside, NJ: Lawrence Erlbaum Associates, 1986.

Becker, Ellen A. and Chere C. Gibson. "Fishbein and Ajzen's Theory of Reason Action: Accurate Prediction of Behavioral Intentions for Enrolling in Distance Education Courses," <u>Adult Education Quarterly, 49-1:</u> 43-56 (1998).

Bevan, Nigel. "Quality in Use: Incorporating Human Tactics into the Software Engineering Lifecycle," in <u>Proceedings of the 3rd International Symposium and Forum on Software Engineering Standards</u>. Whislter, CAN: 1997.

Bevan, Nigel, Jurek Kirakowski, and Jonathan Maissel. "What is Usability?," in <u>Proceedings of the 4th International Conference on HCI</u>. Stuttgart, GE: 1991.

Brynjolfsson, E. "The Productivity Paradox of Information Technology,"
Communications of the ACM, 36-12: 67-77 (1993).

Brynjolfsson, E. and L. Hitt. New Evidence on the Returns to Information Systems.
Cambridge, MA: MIT Press, 1993.

Card, Stuart K., Thomas P. Moran and Allen Newell. The Psychology of Human
Computer Interaction. Mahwah, NJ: Lawrence Erlbaum Associates, 1983.

Chang, Younhwa and Leslie D. Burns. "Attitudinal Versus Normative Influence in the
Purchase of Brand-Name Casual Apparel," Family and Consumer Sciences
Research Journal, 25-1: 79-110 (1996).

Chapanis, Alphonse. "Evaluating Usability," in Human Factors for Informatics
Usability. Brian Shackel and Simon J. Richardson. Cambridge: Cambridge
University Press, 1991.

Chau, Patrick Y. K. "An Empirical Assessment of a Modified Technology Acceptance
Model," Journal of Management Information Systems, 13-2: 185-204 (1996).

Chin, John P., Virginia A. Diehl, and Kent L. Norman. "Development of an Instrument
Measuring User Satisfaction of the Human-Computer Interface," Proceedings of
the ACM CHI 1988 Conference. 213-218. Washingtion, DC: ACM Press (1988).

Currie, Raymond F. and Tim Aubry. "Predicting Intentions of Community Residents
Towards Neighbors with Psychiatric Disabilities," Psychosocial Rehabilitation
Journal, 18-3: 51-67 (1995).

Davis, Fred D. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of
Information Technology," MIS Quarterly, 13-3: 318-337 (1989).

Davis, Fred D. "User Acceptance of Information Technology: System characteristics,
User Perceptions and Behavioral Impacts," International Journal of Man-Machine
Studies, 38: 457-487 (1993).

Davis, Fred D., Richard P. Bagozzi and Paul R. Warshaw. "User Acceptance of
Computer Technology: A Comparison of Two Theoretical Models," Management
Science, 35-8: 982-1003 (August 1989).

Davis, Fred D. and V. Venkatesh. "Measuring User Acceptance of Emerging Information
Technologies: An Assessment of Possible Method Biases," Proceedings of the
28th Annual Hawaii International Conference on System Sciences: 729-736
(1995).

Davis, Fred D. and Paul R. Warshaw. "Choice Sets and Choice Intentions," Journal of Social Psychology, 131-6: 823-831 (1991).

Dennis, Nadine, A., 2nd Lt, USAF. Assessment of Dragon NaturallySpeaking, A Voice Recognition Product. Scott AFB, IL: Air Force Communications Agency Directorate of Technology, 1998.

Dillon, Andrew. "Knowledge Acquisition and Conceptual Models: A Cognitive Analysis of the Interface," in People and Computers III. D. Draper and R. Winder. Cambridge: Cambridge University Press, 1987.

Dillon, Andrew and Michael Morris. "From 'Can They?' to 'Will They': Extending Usability to Accommodate Acceptance Predictions," in Proceedings of the 4th Annual Association for Information Systems, Americas Conference. Baltimore, MD: August 1998.

Dix, A., J. Finlay, and R. Beale. Human-Computer Interaction. Upper Saddle River, NJ: Prentice Hall, 1993.

Dix, Alan J., Janet E. Finlay, Gregory D. Abowd, and Russell Beale. Human-Computer Interaction. London: Prentice Hall Europe, 1998.

Doll, William J, Anthony Hendrickson and Xiaodong Deng. "Using Davis's Perceived Usefulness and Ease-of-use Instruments for Decision Making: A Confirmatory and Multigroup Invariance Analysis," Decision Science, 29-4: 839-869 (1998).

Franke, R. H. "Technological Revolution and Productivity Decline: Computer Introduction in the Financial Industry." Technology Forecasting and Social Change, 31: 143-154.

Fishbein, M. and I. Ajzen. Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Addison Wesley: Reading, MA: 1975.

Gardner, Arthur. "An Approach to Formalized Procedures for User-Centered Design," in Human Factors for Informatics Usability. Brian Shackel and Simon J. Richardson. Cambridge: Cambridge University Press, 1991.

Gould, John D. and Clayton Lewis. "Designing for Usability: Key Principles and What Designers Think," in Readings in Human-Computer Interaction: A Multidisciplinary Approach. Ronald M. Baecker and William A. S. Buxton. Los Altos, CA: Morgan Kaufmann, 1987.

Hartwick, Jon and Henri Barki. "Explaining the Role of User Participation in Information System Use," Management Science, 40-4: 440-465 (1994).

Horn, Pual M. "Perspective Driving Innovation." Except from electronic publication, n. pag. http://www.research.ibm.com/resources/magazines/1998/issue_3/perspective398. html. 16 June 1999.

Igbaria, Magid and Tor Guimaraes. "Testing the Determinants of Microcomputer Usage Via a Structural Equation Model," Journal of Management Information Systems, 11-4: 87-115 (1995).

Intel. "Processor Hall of Fame." Excerpt from unpublished article, n. pag. http://www.intel.com.tw/intel/museum/25anniv/hof/moore.htm. 22 July 1999.

Karat, Clare-Marie, and Christine Halverson, Daniel Horn, and John Karat. "Patterns of Entry and Correction in Large Vocabulary Continuous Speech Recognition Systems," Proceedings of the CHI 99 Conference. 568-575. Pittsburgh, PA: ACM Press, 1999.

Kirakowski, Jurek. and M. Corbett. "SUMI: The Software Usability Measurement Inventory," British Journal of Education Technology, 23-3: 210-212 (1993).

Koslowsky, Meni. "A Comparison of Two Attitude-Behavior Models for Predicting Attrition in Higher Education," Journal of Applied Behavioral Science, 29-3: 359-366 (1993).

Koslowsky, Meni and Michael Hoffman. "Predicting Behaviors on a Computer from Intentions, Attitudes, and Experience," Current Psychology, 9-1: 75-84 (1990).

Landauer, Thomas K. The Trouble with Computers. Cambridge, MA: MIT Press, 1997.

Laudon, Kenneth C., and Jane P. Laudon. Essentials of Management information Systems: Organization and Technology. Upper Saddle River, NJ: Prentice Hall, 1997.

Lee, Chol and Robert T. Green. "Cross-Cultural Examination of the Fishbein Behavioral Intentions Model," Journal of International Business Studies, 22-2: 289-306 (1991).

Lewis, Clayton and John Rieman. "Getting to Know Users and Their Tasks," in Readings in Human-Computer Interaction: Toward the Year 2000. Baecker, Ronald M., Jonathan Grudin, William A. S. Buxton, and Saul Greenberg. San Francisco, CA: Morgan Kaufmann, 1995.

Lewis, James R. "IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instruction for Use," International Journal of Human-Computer Interaction, 7-1: 57-78 (1995).

119

Loveman, G. W. The Productivity of information Technology Capital: An Econmic Analysis. Cambridge, MA: MIT Press, 1986.

Loveman, G. W. An Assessment of the Productivity Impact of information Technology. Cambridge, MA: MIT Press, 1990.

Mantei, Marilyn M. and Toby. J. Teorey. "Cost/Benefit Analysis for Incorporating Human Factors in the Software Lifecycle," Communications of the ACM, 31-4: 428-439 (1988).

Markowitz, Judith A. Using Speech Recognition. Upper Saddle River, NJ: Prentice Hall, (1996).

Markus, Lynne M. "Power, Politics, and MIS Implementation," in Readings in Human-Computer Interaction: A Multidisciplinary Approach. Ronald M. Baecker and William A. S. Buxton. Los Altos, CA: Morgan Kaufmann Publishers, Inc., 1987.

Mathieson, K. "Predicting User Intentions: Comparing the Technology Acceptance Model with the Theory of Planned Behavior," Information Systems Research, 2-3: 173-191 (1991).

Mayhew, Deborah J. The Usability Engineering Lifecycle. San Francisco, CA: Morgan Kaufmann Publishers, Inc., 1999.

McClave, James T, George P. Benson, and Terry Sincich. Statistics for Business and Economics, 7th Edition. Upper Saddle River, NJ: Prentice Hall, 1998.

Molich, R. and J. Nielson. "Improving a Human-Computer Dialog," Communications of the ACM, 33-3: 338-348 (1990).

Moore, Gary C. and Izak Benbasat. "Development of an Instrument to Measure the Perceptions for Adopting an Information Technology Innovation," Information Systems Research, 2-3: 192-222 (1991).

Morris, John. "Software Sneak Preview: Office Suites," PC Magazine, 17-18: 116-127 (October 20, 1998).

Morris, Michael G. and Andrew Dillon. "How User Perceptions Influence Software Use," IEEE Software: 58-64 (July/August 1997).

Nielson, Jakob. Usability Engineering. San Francisco: Morgan Kaufmann Publishers, Inc., 1993.

Nielson, Jakob and Jonathan Levy. "Measuring Usability: Preference vs. Performance," Communications of the ACM, 37-4: 66-75 (1994).

Nielson, Jakob and Robert L. Mack. Usability Inspection Methods. New York: John Wiley & Sons, Inc., 1994.

Nielson, Jakob and V. L. Phillips. "Estimating the Relative Usability of Two Interfaces: Heuristic, Formal, and Empirical Methods Compared," Proceedings of the ACM INTERCHI'93 Conference. 214-221. Amsterdam, the Netherlands: ACM Press (1993).

Norman, Donald A. and Stephen W. Draper. User Centered System Design: New Perspectives in Human-Computer Interaction. Hillside, NJ: Lawrence Erlbaum Associates, 1986.

Poor, Alfred. "Watch What You Say," PC Magazine, 17-5: 129-148 (10 March 1998).

Pressman, Roger S. Software Engineering: A Practitioner's Approach. New York: McGraw-Hill, Inc., 1992.

Rogers, Everett M. Diffusion of Innovation. New York: Free Press, 1983.

Shackel, Brian and Simon J. Richardson. Human Factors for Informatics Usability. Cambridge: Cambridge University Press, 1991.

Shneiderman, Ben. Designing the User Interface. Reading, Massachusetts: Addison Wesley, 1998.

Smith, Wanda, J. ISO and ANSI Ergonomic Standards for Computer Products. Upper Saddle River, NJ: Prentice Hall, 1996.

Szajna, Bernadette. "Software Evaluation and Choice: Predictive Validation of the Technology Acceptance Instrument," MIS Quarterly, 18-3: 319-324 (1994).

Taylor, Shirley and Peter A. Todd. "Understanding Information Technology Usage: A Test of Competing Models," Information Systems Research, 6-2: 144-176 (1995).

Toft, Dorte. "Who Needs Speech Recognition?," IDG News Service, n. pag. http://www2.pcworld.com/pcwtoday/article/0,1510,11324,00.html. 10 June 1999.

Ulrich, John. "Counselling of Alcohol-Dependents Based on a Model of Change to Abstinence," Counselling Psychology Quarterly, 3-2: 183-190 (1990).

Venkatesh, Viswanath. "Creation of Favorable User Perceptions: Exploring the Role of Intrinsic Motivation," MIS Quarterly, 23-2: 239-260 (1999).

Venkatesh, Viswanath and Fred D. Davis. "A Model of the Antecedents of Perceived Ease of Use: Development and Test," Decision Sciences, 27-3: 451-478 (1996).

Welsh, W. N. and A. Gordon. "Cognitive Mediators of Aggression," <u>Criminal Justice and Behavior, 18-2:</u> 125-146 (1991).

Whiteside, John, John Bennett, and Karen Holtzblatt. "Usability Engineering: Our Experience and Evolution," <u>Handbook of Human-Computer Interaction</u>. North Holland, Amsterdam: Elsevier Science Publishers B.V., 1988.

Vita

Captain Alan P. Fiorello was born on 21 February 1965 in Los Angeles, California. He graduated from Santa Monica High School in June 1983. Captain Fiorello enlisted in the Air Force in December 1983 and served as an avionics technician for the F-111A and EF-111A aircraft at Mountain Home AFB, Idaho, until December 1987. In February 1988, he enrolled at San Diego State University and was accepted in the AFROTC program in 1989. He graduated on December 1990 with a Bachelor of Art Degree in Industrial Arts and was commissioned in May 1991.

Upon Completion of Basic Communications and Computer Systems Officer Training at Keesler AFB, Mississippi, Captain Fiorello was assigned to the 4th Space Operations Squadron, Falcon AFB, Colorado, working current operations of the MILSTAR satellite. Next he was assigned to Langley AFB, Virginia, where he worked in the HQ ACC Inspector General Directorate. In May of 1998, he entered the Graduate Information Resource Management Degree Program at the Air Force Institute of Technology at Wright-Patterson AFB, Ohio. After graduation, Captain Fiorello will be assigned to the HQ AFSOC Communications and Information Directorate at Hurlburt Field, Florida.

Permanent Address: 4725 Williamsburg Glade
                    Williamsburg, VA. 23185

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1999	Master's Thesis

4. TITLE AND SUBTITLE

Power, Performance, and Perception (P3): Integrating Usability Metrics and Technology Acceptance Determinants to Validate a New Model for Predicting System Usage

5. FUNDING NUMBERS

6. AUTHOR(S)

ALAN P. FIORELLO, Captain, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
2950 P. Street
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GIR/LAS/99D-3

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFRL/HECA
Mr. David Williamson
148 8th St.
Wright-Patterson AFB, OH 45433

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Michael G. Morris, Major, USAF
(937) 255-3636 x4578
michael.morris@afit.af.mil

12a. DISTRIBUTION AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Currently, there are two distinct approaches to assist information technology managers in the successful implementation of office automation software. The first approach resides within the field of usability engineering, while the second approach is derived from the discipline of management information systems (MIS). However, neither approach has successfully produced conclusive evidence that explains what characteristics facilitate system use as well as influence user acceptance of the system.

This study reports on the validity of a new model, entitled the Power, Performance, Perception (P3) model, that links the constructs of usability engineering to user acceptance. Additionally, speech recognition software (SRS) was used in an experimental setting to validate the P3 model. This research also examined the viability of employing SRS in an Air Force office environment.

The results of this study failed to validate the P3 model. However, an alternate model for predicting user acceptance, the Usability-Acceptance Model, did emerge from the research which showed that the usability metric of user satisfaction can explain 53% of the variance of user intention to use a new technology. Additionally, the results of this study indicate that users in a typical Air Force office environment would utilize SRS for text processing.

14. SUBJECT TERMS

Technology Assessment, Usability, Speech Recognition, Man Computer Interface, Productivity

15. NUMBER OF PAGES

132

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL